Biostatistical Methods I (P8130)
Hw2 (10/7/2019)
Hana Akbarnejad (ha2546)
R file attached

Problem 1 – a)

P(D) = 5 x 10^{-6} per year n(NY) = 8.5 x 10^{6}
P(W) = 6.02 x 10^{-6} per year n(W) = 0.428 x 8.5 x 10^{6} = 3638000
P(A) = 0.39 x 10^{-6} per year n(A) = 0.14 x 8.5 x 10^{6} = 1190000
P(B) = 0.31 x 10^{-6} per year n(B) = 0.243 x 8.5 x 10^{6} = 2065500

$$P(D = 30) = \binom{n}{k} p^k q^{n-k} = \binom{8500000}{30} 0.000005^{30}(1 - 0.000005)^{8500000-30} = 0.00934$$

Answer checked with the function dbinom() in R (dbinom(30, 8500000, 0.000005)).
Also, because P is very small and n is very large, we can use Poisson approximation to Binomial:

$$\lambda = \text{np} = 8.6 * 10^6 * 5 * 10^{-6} = 42.5$$

$$P(D = 30) = Poisson(\lambda) = \frac{\lambda^n e^{-\lambda}}{n!} = \frac{42.5^{30} e^{-42.5}}{30!} \cong 0.00934$$

This result also checked with the function dpois() in R (dpois(30, 42.5)).

Problem1-b)

$$P(D = 30\ in\ W) = \binom{3638000}{30} 0.00000602^{30}(1 - 0.00000602)^{3638000-30} = 0.01899915$$

$$Poisson\ approximation: \lambda = \text{np} = 21.9, \qquad P(D = 30\ in\ W)\ \frac{21.9^{30} e^{-21.9}}{30!} \cong 0.19$$

$$P(D = 30\ in\ A) = \binom{1190000}{30} 0.0000039^{30}(1 - 0.0000039)^{1190000-30} \cong 0$$

$$Poisson\ approximation: \lambda = \text{np} = 0.464, \qquad P(D = 30\ in\ A)\ \frac{0.464^{30} e^{-0.464}}{30!} \cong 0$$

$$P(D = 30\ in\ B) = \binom{2065500}{30} 0.0000031^{30}(1 - 0.0000031)^{2065500-30} \cong 0$$

$$Poisson\ approximation: \lambda = \text{np} = 0.64, \qquad P(D = 30\ in\ B)\ \frac{0.64^{30} e^{-0.64}}{30!} \cong 0$$

Problem 2 -a)

In this part we are interested in the absolute changes of BMI for baseline and after 6 months follow-up. We investigate within each of the Control and Intervention groups. For this analysis we need paired t-test because we are comparing one group's situation in

different times, so we cannot assume independence. To do this, First I cleaned up the dataset so that I only have pre and post bmi columns and also added a variable which shows the difference between post-treatment and pre-treatment values. Our null hypothesis is that there is no significant difference between pre and post groups of each of control and intervention groups (H_0: $\mu_1 = \mu_2$ or $\Delta = 0$), and our alternative hypothesis is that the mean of differences vary significantly (H_1: $\mu_1 \neq \mu_2$ or $\Delta \neq 0$).

For investigating within intervention group, I computed the mean and standard deviation of the difference of these two subgroups, and computed the t-statistics using the formula:

$$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{-0.7611111}{1.438705 / \sqrt{36}} = -3.17415$$

With the critical value ($t_{n-1, 1-\alpha/2}$) = $t_{35, 0.975}$ = 2.030108. According to our decision rule which is rejecting H_0 if $|t| > t_{n-1, 1-\alpha/2}$ and we fail to rejecting H_0 if $|t| < t_{n-1, 1-\alpha/2}$. Because$|-3.17415| > 2.030108$, we reject H_0 and conclude that there seems to be a statistically significant difference between pre and post groups of the intervention group. These calculations have been done in R and also checked with t.test() function. The P-value for this test is 0.003127 which means that the probability of observing a result this extreme or more extreme is 0.3% and since it is smaller than our significance level (0.05) we reject null hypothesis.

For obtaining the 95% confidence interval:

$$\bar{d} - t_{n-1, 1-\alpha/2} s_d / \sqrt{n} \leq \Delta \leq \bar{d} + t_{n-1, 1-\alpha/2} s_d / \sqrt{n}$$

$$-0.7611111 - (2.030108)\left(1.438705 / \sqrt{36}\right) \leq \Delta \leq -0.7611111 + (2.030108)\left(1.438705 / \sqrt{36}\right)$$

$$-1.2478990 \leq \Delta \leq -0.2743233$$

The 95% confidence interval is (-1.2478990 , -0.2743233) which means we are 95% sure that the true mean of difference in the population lies somewhere between -0.2743233 and -1.2478990.

I then performed the same test for the control group with the similar hypothesis and process:

$t = \frac{\bar{d}-0}{s_d / \sqrt{n}} = \frac{0.2833333}{0.9676186 / \sqrt{36}} = 1.7569$ with the critical value ($t_{n-1, 1-\alpha/2}$) = $t_{35, 0.975}$ = 2.030108.

since $|t| < t_{n-1, 1-\alpha/2}$, we fail to reject H_0 and conclude that there is no significant difference between means in pre and post groups of the control group. The P-value for this test is 0.08768 (the probability of observing as extreme or more extreme) which is greater than 0.05 and confirms this result. The confidence intervals for this test is:

$$\bar{d} - t_{n-1, 1-\alpha/2} s_d / \sqrt{n} \leq \Delta \leq \bar{d} + t_{n-1, 1-\alpha/2} s_d / \sqrt{n}$$

$0.2833333 - (2.030108)(\,^{0.9676186}/_{\sqrt{36}}) \leq \Delta \leq 0.2833333 +$
$(2.030108)(\,^{0.9676186}/_{\sqrt{36}})$
$-0.04406169 \leq \Delta \leq 0.61072836$ which means that the true difference if the means of pre and post in the control population lies between these two values.

Problem 2 – b)

In this part, we are taking the differences between pre and post groups in control, and in intervention, and we are comparing these two groups of differences. We can consider these two groups as independent groups and perform independent t-test to compare the means.

For this test, we should first test the equality of variances. The test statistics would be:

$$F = \frac{{s_1}^2}{{s_2}^2} \sim F_{n1-1,n2-1}$$

$$F = \frac{0.9676186^2}{1.438705^2} = 0.45234$$

Our testing hypothesis is H_0: $\sigma_1^2 = \sigma_2^2$ and H_1: $\sigma_1^2 \neq \sigma_2^2$.
We reject H_0 if $F > F_{n1-1,n2-1,1-\alpha/2}$ or $F < F_{n1-1,n2-1,\alpha/2}$

We fail to reject H_0 if $F_{n1-1,n2-1,\alpha/2} \leq F \leq F_{n1-1,n2-1,1-\alpha/2}$

Computing our test statistics, our F equals 0.45234 and our critical values equal 0.5099207 and 1.961089. since $F < F_{n1-1,n2-1,\alpha/2}$ $(0.45234 < 0.5099207)$, we reject null hypothesis and conclude that there is significant difference between variances of these two groups. Since $F < 1$, the P-value will be: $2 * P(F_{n1-1,n2-1} < F \mid H_0) = 0.02141$ which is smaller than our significance level and we can reject null.

Now that I understood that these two groups have unequal variances, I use the formula for two-sample independent t-test with unequal variances for t-statistics, to test the null hypothesis that the true means are equal and alternative hypothesis that means differ significantly. For computing the test statistics:

$$t = \frac{X_1 - X_2}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} = \frac{0.2833333 - (-0.7611111)}{\sqrt{\frac{0.9676186^2}{36} + \frac{1.438705^2}{36}}} = 3.6144$$

We also need the degree of freedom:

$$df' = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{0.08350438422^2}{\frac{(\frac{0.9676186^2}{36})^2}{35} + \frac{(\frac{1.438705^2}{36})^2}{35}} = 61.285, \text{ and } df'' = 61$$

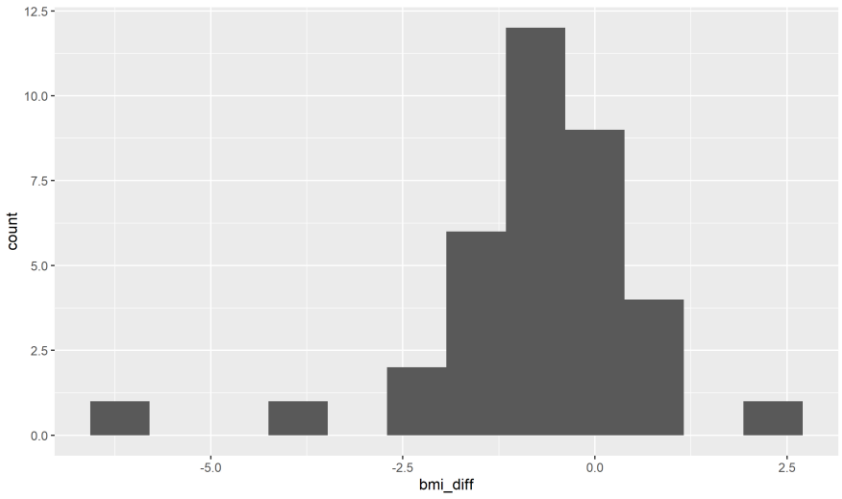
We reject null hypothesis if $|t| > t_{df",1-\alpha/2}$ and fail to reject if $|t| < t_{df",1-\alpha/2}$. Here, $|t| > t_{df",1-\alpha/2}$ (1.999624) and our p-value is 0.0006093 which is smaller than 0.05, so we reject null hypothesis and conclude that the absolute changes of pre and post groups are significantly different between control and intervention groups. These results imply that the exercise program that the intervention group received was statistically effective to the changes of BMI in our sample!

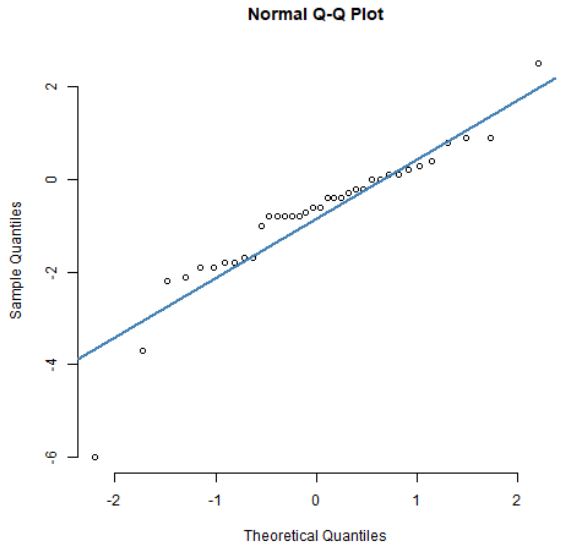
Problem 2 – c - i)

For part a, we assume normality and for part b we assume normality and independence of samples.

To check the normality of these cases, I created histograms for the differences between pre and post groups of our intervention and our control groups. I also used QQ-plots to see any deviation from the normal line that might not be easy to be detected using histograms.

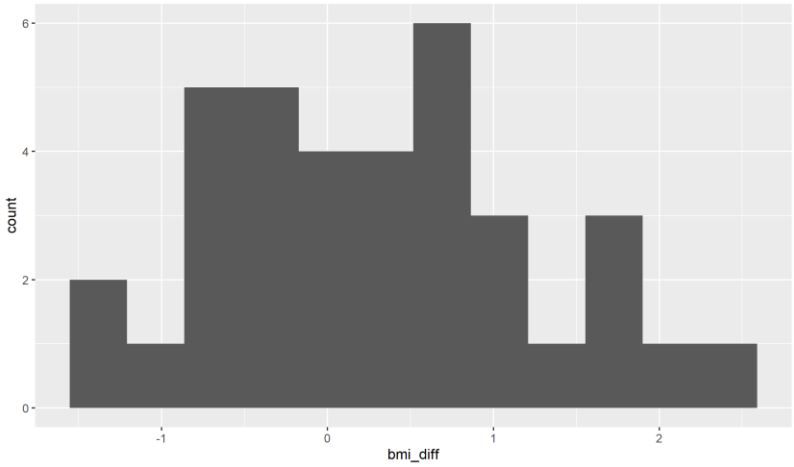
The histogram and QQ-plot for the intervention group:

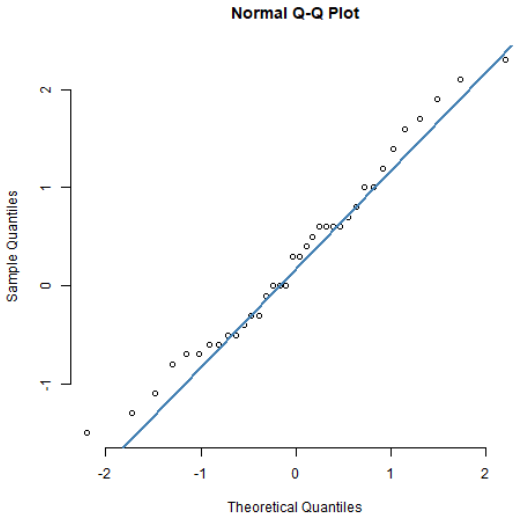




As we can observe this group has obvious deviation from normality.

Doing the same procedure for the control group, we will have:





For these plots we also observe deviation from normality and in particular, tails seem to be much heavier than a normal curve.

2 – c – ii)

Some statistical tests are robust to the normality assumption, such as ANOVA, one sample, two-sample, and paired t-test, Regression, and Design of Experiment (DOE). So, although normality is underlying assumption of these tests, their results are most of the times reliable regardless of the normality of samples. However, the sample size is very important and if our sample is not large enough, our results would deviate from what is truly going on in the population. In such situations, one way is to perform simulations with different available software and use random number generators. In rare cases, using non-parametric analysis might be useful as well.

Problem 3)

For the exact case, P(Close) = P(C) = 0.6
P(Not close) = q(C) = 1 – P(C) = 0.4
We can see that X~Binom(20, 0.6), and we want to know tha $P(X \geq 0.5)$.

$$P(X \geq 10) = 1 - P(X < 10) = 1 - (x+a)^n = 1 - \sum_{x=0}^{9} \binom{20}{x} 0.6^x \, 0.4^{20-x} \cong 0.8725$$

Also checked with R “pbinom” function.

For approximation, we can observe that our n is small (n = 20 $<$ 100) and P is large ($P = 0.6 > 0.01$), so probably we cannot obtain a good approximation using Poisson. However, I compute that to observe the difference:
λ = np = 20 x 0.6 = 12, so X~ Poi (12).

$$P(X \geq 10) = 1 - P(X < 10) = 1 - (x + a)^n = 1 - \sum_{x=0}^{9} \frac{12^x \, e^{-12}}{x\,!} \cong 0.7576$$

Also checked with the R function "ppois".
So, we may think that the Poisson distribution here is not a good approximation to Binomial distribution, because P is not small enough and n is not large enough.

Problem 4 - a)

In this problem, we need to know if there is any significant difference between the effectiveness of drug 1 compared to drug 2 on duration of sleep of students. Also, with the directionality implied, we can change the question to: whether drug 2 was more effective than drug 1. In this problem, I used R for performing and checking the calculations when possible.

So, if we consider μ_1 as the mean increase(difference) in sleep hours in population as a result of taking drug 1 and μ_2 as the mean increase in sleep hours in population as a result of taking drug 2, our null hypothesis would be that there is no significant different between means of sleep hours change in these two groups (H_0: $\mu_1 = \mu_2$), and our alternative hypothesis as the mean of increase in amount of sleep in one of these two groups has been statistically significantly greater than another group (H_1: $\mu_2 > \mu_1$ or equally $\mu_2 - \mu_1 > 0$).

Because we are trying to compare two groups here, we should use paired t-test, and because we care about directionality in this question, we use one-sided paired t-test. So, we need to have the difference between the amount of sleep after using drug 1 and drug 2 for each of the 10 students (d), compute $\bar{d}$ as its mean, and s_d as standard deviation of the difference. So, our test-statistics would be:

$$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{1.58}{1.23 / \sqrt{10}} = 4.062$$

Now, we should set the decision rule. We reject H_0 if $|t| > t_{n-1,\,1-\alpha}$ and we fail to reject H_0 if $|t| < t_{n-1,\,1-\alpha}$ (we use $1 - \alpha$ since it is one-sided analysis). To compute the critical value:

$t_{n-1,\,1-\alpha} = t_{9,\,0.95} = 1.833$

Based on $|t| > t_{n-1,\,1-\alpha}$ ($4.062 > 1.833$), at 0.05 confidence level, we reject H_0, and conclude that there is a statically significance difference between the true effect of these two drugs on sleep duration of students. Because we used directionality in computations, we can say that it seems that drug 2 has been significantly more effective in increasing sleep hours in students compared to drug 1. Also, the p-value here is 0.002833 which implies that we have strong evidence against the null (i.e. the probability of observing something as extreme or more extreme, when null hypothesis true ($\mu_1 = \mu_2$) is around 0.28%) which is smaller than α (significance level), so we reject null hypothesis.

4 – b)

Without directionality, the 95% confidence interval for the mean increase(difference) of hours of sleep would be:

$$\bar{d} - t_{n-1,\,1-\alpha}\, {}^{s_d}/_{\sqrt{n}} \leq \Delta \leq \bar{d} + t_{n-1,\,1-\alpha}\, {}^{s_d}/_{\sqrt{n}}$$

$$1.58 - (1.833)\left({}^{1.23}/_{\sqrt{10}}\right) \leq \Delta \leq 1.58 + (1.833)\left({}^{1.23}/_{\sqrt{10}}\right)$$

$0.867 \leq \Delta \leq 2.293$

This would means that we are 95% confident that the true population mean of the increase in sleeping hours lies between 0.867 hours increase and 2.293 hours increase.

However, because of the directionality implied in this question, we can have $0.867 \leq \Delta$ as lower one-sided confidence interval and $\Delta \leq 2.293$ as upper one-sided confidence interval. Here, we accept the lower bound CI meaning that we are 95% confident that the true difference in the population is greater than 2.293

4 – c - i)

Effect size equals 2.293 – 0.867 = 1.426

$$\text{Power} = \Phi\left(z_\alpha + \frac{\Delta}{\frac{\sigma_d}{\sqrt{n}}}\right) = \Phi\left(-z_{1-\alpha} + \frac{\Delta}{\frac{\sigma_d}{\sqrt{n}}}\right) = \Phi\left(-1.645 + \frac{1.426}{\frac{1.229995}{\sqrt{10}}}\right) = \Phi(2.0262) \cong 98\%$$

This means that there is 98% probability that we reject null hypothesis when we are 95% confident that the true difference in the hours of sleep is between 0.867 and 2.293 hours.

4 – c – ii)

Power is the conditional probability that we will reject the null hypothesis given that the null hypothesis is not true by a specified amount and given certain other specifications, such as sample size and significance level. In Posteriori/Observed/Post-hoc power analysis, we find how much power we would have based on the effect size estimate from our data (if we had a specified number of cases), and not based on the true effect size we are examining. Posteriori power advocates argue that at the observed effect size with a high power computed, there is evidence for the null hypothesis being true although the result is not statistically significant. However, "power approach paradox" (PAP) says that higher observed power does not imply stronger evidence for a null hypothesis that is not rejected. The problem with post-hoc power analysis is that we use estimated effect size and using estimated effect size overestimates the power drastically. Additionally, when we have performed our analysis and have observed the p-value, calculating the power is mathematically redundant and adds no value to interpretations in research studies because of the one-to-one relationship between p-value and observed power (so, nonsignificant p values always correspond to low observed power). So, although statistical power

calculation is a valuable tool in planning a study, if used as aid in the interpretation of the experimental results it can work as a form of confirmation bias.

Reference:

Hoenig, J. M. , Heisey, D. M., The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis. The American Statistician. February 2001, Vol. 55, No. 1