STAT 514 Homework 1

Due Sep 4

1. Randomization Test: In a study of egg cell maturation, the eggs from each of four female frogs were divided into two batches and one batch was exposed to progesterone. After two minutes, the cAMP content was measured. It is believed that cAMP is a substance that can mediate cellular response to hormones.

Frog	Control	progesterone	diff
1	6	4	2
2	4	5	-1
3	5	2	3
4	4	2	2

Please use randomization test to test whether there is mean difference between control group and progesterone. (Hint: it is a paired t-test. Under the Null hypothesis, within each pair, treatment control and progesterone are exchangable.)

Solution:

For each pair, we can flip control and progesterone, for example

Frog	Control	progesterone	diff
1	4	6	-2
2	4	5	-1
3	5	2	3
4	4	2	2

Totally, we have 2^4 different data sets, which are homogeneous if Null if true (No difference between control and progesterone). For each generated data set, we compute the sample mean difference, the 16 differenct sample mean difference are {-2, -0.5,-1, 0.5,-1, 0.5, 0, 1.5, -1.5, 0, -0.5, 1, -0.5, 1, 0.5, 2}, where 1.5 is the sample mean difference of the obeserved data set.

Among the 16 values, 25% of them have absoluted value larger or equal 1.5. (That is the pvalue for two side test).

Among the 16 values, 12.5% of them have absoluted value larger or equal 1.5. (That is the pvalue for one side test).

2. Suppose we have 10 random variable generated in the following manner. I first generate a variable X from rom a Normal distribution with mean 5 and variance 1. I then randomly sample 10 variables Y_1 to Y_{10} from a Normal distribution with mean 0 and variance .5 and add the variable to each one. In other words, this 10 random variables follows

$$Z_i = X + Y_i, \quad i = 1, \ldots, 10,$$

where $X \sim N(5, 1)$, Y_i independently $\sim N(0, 0.5)$.

Please figure out $E(Z_i)$, $Var(Z_i)$, $Cov(Z_i, Z_j)$ $(i \neq j)$, $Var(\bar{Z})$ $(\bar{Z} = \sum Z_i/10)$.

Solution

$$
\begin{aligned}
E(Z_i) &= E(X+Y_i) = E(X) + E(Y_i) = 5 + 0 = 5; \\
Var(Z_i) &= Var(X+Y_i) = Var(X) + Var(Y_i) = 1 + 0.5 = 1.5; \\
Cov(Z_i, Z_j) &= Cov(X+Y_i, X+Y_j) = Cov(X,X) + Cov(X,Y_j) \\
&\quad + Cov(Y_i, X) + Cov(Y_i, Y_j) = Cov(X,X) = Var(X) = 1; \\
Var(\bar{Z}) &= Var(X+\bar{Y}) = Var(X) + Var(\bar{Y}) = Var(X) + Var(Y_i)/10 = 1.05;
\end{aligned}
$$

3. Let us continue to study the keyboard experiement in the first lecture. Let y be the amout of time used to type up a manuscript. Note that y depends on keyboard, manuscript, whether the manuscript has already been typed, and experimental error. Let μ_A and μ_B denote the effects of keyboard A and B respectively, τ_i the effect of manuscript i for $i = 1, 2, 3, 4, 5, 6$ and ϵ denotes the experimental error. Let α_l denote the learning effect. We are interested in estimating the difference between μ_B and μ_A. Suppose Design 2 from the lecture notes is used in the experiment, which is

$$1.A - B; 2.B - A; 3.A - B; 4.B - A; 5.A - B; 6.A - B.$$

The statistical model for the amount of time used in $1 : A$, denotes by y_{1A} is

$$y_{1A} = \gamma + \mu_A + \tau_1 + \epsilon_{1A},$$

and the model for the amount of time used in $1 : B$ is

$$y_{1B} = \gamma + \mu_B + \tau_1 + \alpha_l + \epsilon_{1B},$$

where γ is some constant value.

(a) Is α_l positive or negative? Why it is not included in first model?
Solution: Negative, since learning effect occurs only when the typist types the manuscript the second time.

(b) Write down the statistical models for the other runs.
Solution: Omitted.

(c) Regardless of which design is used, what is the *simplest* estimate for $\mu_A - \mu_B$, using $y_{1A}, y_{1B}, \ldots, y_{6A}, y_{6B}$
Solution: $\hat{\mu}_A = \sum_i y_{iA}/6$, $\hat{\mu}_B = \sum_i y_{iB}/6$, $\hat{\mu}_A - \hat{\mu}_B = \sum_i (y_{iA} - y_{iB})/6$.

(d) Explain why the third design is the best one. (Hint: consider the biasedness of the estimator of $\mu_A - \mu_B$ under three different designs)
Solution:
Design 1: $E(\hat{\mu}_A - \hat{\mu}_B) = \mu_A - \mu_B - \alpha_l$;
Design 2: $E(\hat{\mu}_A - \hat{\mu}_B) = \mu_A - \mu_B - \alpha_l/3$;
Design 3: $E(\hat{\mu}_A - \hat{\mu}_B) = \mu_A - \mu_B$.
Thus, under the design 3, the estimator is unbiased.