STAT 514 Homework 2

Due: Sep 11

1. In hypothesis testing, if I keep n fixed and decrease the significance level ($\alpha \to 0$), what happens to the Type I and Type II error probabilities? If I keep α fixed and increase sample size n, what happens to the Type I and Type II error probabilities? In order to make both Type I AND Type II error decrease, what should we do?

   solution: Type I error decrease, Type II increase; Type I error is fixed, and Type II decrease; We have to increase sample size and decrease α at the same time.

2. You are requested to design an experiment to compare the typing efficiency of three different types of keyboards denoted by A, B and C. Two typists, denoted by T_1 and T_2, are employed and six standard manuscripts $m_1, \ldots, m_6$ are used. Please propose your experimental plan.

   solution: Both typists and manuscripts should be treated as blocks. In all, we have 12 blocks. Within each block, all treatments (three types of keyboards) should be applied. The order of using the keyboards should be randomized. If we want to eliminate the learning effects entirely, balanced randomization should be used. Advantages of randomization and blocking refer to the notes.

3. A vendor submits lots of fabric to a textile manufacturer. The manufacturer wants to know if the lot average breaking strength exceeds 200 psi. if so, she wants to accept the lot. Past experience indicates that a reasonable value for the variance of breaking strength is 100 (psi)^2. The hypotheses to be tested are

   $$H_0 : \mu = 200 \text{ vs } H_1 : \mu > 200.$$

   The manufacturer decides to randomly select a number of specimens, measure their breaking strengths and test the hypotheses with = 5%. Suppose she wants to guarantee that average breaking strength 210 or higher should be detected with probability at least 95%. What is the minimum number of specimens she should check? (Note that the test statistic under H_1 is also normal distribution, thus you can directly compute the power of the test $Pr(\text{test statistic} > \text{critical value}|\mu = 210)$, and do not need to generate O.C. curves)

   solution:

   Test statistics $t.s. = \sqrt{n}(\bar{x} - \mu_0)/\sigma$, where $\mu_0 = 200$, $\sigma = 10$. The critical value= 1.645. Thus we need to find out minimum value of n, such that $Pr(\sqrt{n}(\bar{x}-200)/10 > 1.645|\mu = 210)$. Since $x \sim N(210, 10^2)$, $(\bar{x} - 200) \sim N(10, 100/n)$, $\sqrt{n}(\bar{x} - 200)/10 \sim N(\sqrt{n}, 1)$. $Pr(N(\sqrt{n}, 1) > 1.645) = Pr(N(0, 1) > 1.645 - \sqrt{n})$. Since $Pr(N(0, 1) > -1.645) = 0.95$, $\sqrt{n} > 1.645 - (-1.645)$, hence, the minimum value for $Pr(N(\sqrt{n}, 1) > 1.645) = 0.95$ is $n = 11$.

4. Please verify that

$$\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{..})^2 = \sum_{i=1}^{a} n_i(\bar{y}_{i.}-\bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2$$

.

solution:

$$\begin{aligned}
&\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{..})^2 \\
=&\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.}+\bar{y}_{i.}-\bar{y}_{..})^2 \\
=&\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(\bar{y}_{i.}-\bar{y}_{..})^2 + 2\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})(\bar{y}_{i.}-\bar{y}_{..}) \\
=&\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2 + \sum_{i=1}^{a} n_i(\bar{y}_{i.}-\bar{y}_{..})^2 + 2\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})(\bar{y}_{i.}-\bar{y}_{..})
\end{aligned}$$

And

$$\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})(\bar{y}_{i.}-\bar{y}_{..}) = \sum_{i=1}^{a}\{(\bar{y}_{i.}-\bar{y}_{..})\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})\} = \sum_{i=1}^{a}\{(\bar{y}_{i.}-\bar{y}_{..})0\} = 0.$$

5. An article in the *Journal of Strain Analysis* (vol.18, no.2, 1983) compares several procedures for predicting the shear strength for steel plate girders. Data for nine girders in the form of the ratio of predicted to observed load for two of these procedures, the Karlsruhe and Lehigh Methods, are given as follows:

Girder	Karlsruhe Method	Lehigh Method
S1/1	1.186	1.061
S2/1	1.151	0.992
S3/1	1.322	1.063
S4/1	1.339	1.062
S5/1	1.200	1.065
S2/2	1.402	1.178
S2/3	1.365	1.037
S2/4	1.537	1.086
S2/5	1.559	1.052

Please investigate the normality of both two samples (two columns) and the difference in rations of the two methods

Solution: Use QQ plot, histogram or hypothesis test by SAS code. There does not appear to be any significant departure from normality.