IEOR 151 – Homework 1
Due Friday, September 26, 2013 in class

1. For each the following scenarios, would you (i) accept the null hypothesis, (ii) reject the null hypothesis, or (iii) gather additional data and information before making a decision? Explain your reasoning. Note: The numbers in the scenarios below are fictional.

   (a) The null hypothesis is that hospitals using paper records have equal costs compared to hospitals using electronic records, the difference in average costs per patient is \$213 more for hospitals using paper records, and $p = 0.051$. (2 points)

   Collect more data, because the potential savings are substantial if the cost difference is truly as large is \$213. The current p-value is very slightly above $\alpha = 0.05$, and so we should collect more data to better discriminate between outcomes.

   (b) The null hypothesis is that diners sitting at black tables spend as much as diners sitting at dark brown tables, the difference between the average spend amount is 10 cents, and $p = 0.049$. (2 points)

   Accept the null hypothesis, because even though we would reject the hypothesis at the $\alpha = 0.05$ level, the risk in choosing incorrectly is reduced since the average difference is 10 cents.

   (c) The null hypothesis is that there is no botulism bacteria in honey, 10% of store bought honey is found to contain botulism bacteria, and $p = 0.023$. (2 points)

   Reject the null hypothesis, because there is greater risk with accepting the null hypothesis and the p-value is quite a bit lower than the $\alpha = 0.05$ level.

   (d) The null hypothesis is that the speed of light in a vacuum on earth is the same as the speed of light in a vacuum in space, the measured difference is $20,000$ meters per second, and $p = 0.009$. (2 points)

   Accept the null hypothesis, because there is significant past scientific evidence supporting the null hypothesis and so we would only reject if the hypothesis had a very low p-value.

2. Suppose 5 different hypothesis tests have been conducted, with p-values of: Test 1 ($p = 0.07$), Test 2 ($p = 0.002$), Test 3 ($p = 0.011$), Test 4 ($p = 0.003$), Test 5 ($p = 0.04$).

   (a) Using the Bonferroni correction, which tests should be accepted or rejected when the family-wise error rate is $\alpha = 0.05$. (2 points)

   Since there are five tests, the Bonferroni correction states that a hypothesis should be rejected if $p < \alpha/5 = 0.01$. Thus, Tests 2 and 4 should be rejected and Tests 1, 3, and 5 should be accepted.

   (b) Using the Holm-Bonferroni method, which tests should be accepted or rejected when the family-wise error rate is $\alpha = 0.05$. (3 points)

   We begin by arranging the p-values in increasing order: 0.002, 0.003, 0.011, 0.04, 0.07. We need to determine the smallest k such that the k-th p-value in the arranged list is greater than $Q_k = \alpha/(5 + 1k)$. For $k = 1, \ldots, 5$, the rounded values of Q_k are 0.01, 0.0125, 0.017, 0.025, and 0.05. In this case, $k = 4$ is that smallest k. As a result, we reject hypothesis corresponding to the first three p-values in the ordered list and accept the remaining. Thus, Tests 2, 3, and 4 should be rejected and Tests 1 and 5 should be accepted.

3. Suppose $X_i \sim \mathcal{N}(\mu, \sigma^2)$ (for $n = 10$ data points) is iid data drawn from a normal distribution with mean μ and variance $\sigma^2 = 20$. Here, the mean is unknown, and we would like to determine if the mean is $\mu_0 = 0$ (decision d_0) or $\mu_1 = 5$ (decision d_1). Lastly, suppose our loss function is

   - $L(\mu_0, d_0) = 0$ and $L(\mu_0, d_1) = 2$;
   - $L(\mu_1, d_0) = 4$ and $L(\mu_1, d_1) = 0$.

   (a) Suppose γ^* is the threshold for the minimax hypothesis test. Without doing any calculations, what region must γ^* lie in? Explain your reasoning. Hint: There are three possibilities: $\gamma^* \in (0, 5)$, $\gamma^* \in (0, 2.5)$, or $\gamma^* \in (2.5, 5)$. (2 points)

   We must have $\gamma^* \in (0, 2.5)$, because (i) the loss for choosing d_0 when $\mu = \mu_1$ is higher than the loss for choosing d_1 when $\mu = \mu_0$ (and $\mu_1 = 5 > \mu_0 = 0$) and so (ii) we would like to choose d_0 less often.

   (b) Using a binary search and a z-table, compute γ^* to within an accuracy of ± 0.1. (4 points)

   We would like to find γ^* that satisfies

$$2 \cdot (1 - \Phi(\sqrt{10}(\gamma - 0)/\sqrt{20})) = 4 \cdot \Phi(\sqrt{10}(\gamma - 5)/\sqrt{20}),$$

or equivalently

$$2 \cdot (1 - \Phi(0.707 \cdot (\gamma - 0))) = 4 \cdot \Phi(0.707 \cdot (\gamma - 5)).$$

Since $\gamma^* \in (0, 2.5)$, our first guess is half-way: $(0+2.5)/2 = 1.25$. The first guess and subsequent guesses of the binary search are summarized in the table below. The final solution is that $\gamma^* \in (2.27, 2.35)$, which is within accuracy of ± 0.08. For reference, the final solution within an accuracy of ± 0.01 is $\gamma^* \in (2.27, 2.28)$.

Step	γ	LHS	RHS
1	1.25	0.3768	0.0160
2	1.88	0.1838	0.0548
3	2.19	0.1215	0.0939
4	2.35	0.0966	0.1220
5	2.27	0.1085	0.1072

4. Suppose we would like to select the optimal level of inventory of raspberries using a newsvendor model with production costs. Here, demand $X \sim \mathcal{N}(5200, 4000)$ is in units of boxes of raspberries, fixed costs are 1325 dollars, variable costs are 0.27 dollars per box of raspberries, holding costs are 1.50 dollars per box of raspberries, and the sale price is 2.99 dollars per box of raspberries. What is the optimal inventory level? (3 points)

   The optimal inventory level δ^* is given by

   $$F(\delta^*) = \Phi((\delta^* - 5200)/\sqrt{4000}) = \frac{2.99 - 0.27}{2.99 + 1.50} = 0.6058.$$

   From a z-table, we have that $\Phi(0.27) = 0.6064$, and so we would like to choose δ^* such that

   $$(\delta^* - 5200)/\sqrt{4000} = 0.27 \Rightarrow \delta^* = 5217.$$

5. Suppose we would like to select the optimal level of inventory of newspapers using a newsvendor model without production costs. Here, demand $X \sim \mathcal{U}(100, 200)$ is in units of newspapers, holding costs are 2.00 dollars per newspaper, and the sale price is 1.50 dollars per newspaper. What is the optimal inventory level? (4 points)

   The optimal inventory level δ^* is given by

   $$F(\delta^*) = \frac{1.50}{2.00 + 1.50} = 0.4286.$$

   Since we have a uniform distribution, we have that for $\delta^* \in (100, 200)$:

   $$\begin{aligned} F(\delta^*) &= \int_{100}^{\delta^*} \frac{1}{200 - 100} du = 0.4286 \\ &\Rightarrow (\delta^* - 100)/100 = 0.4286 \\ &\Rightarrow \delta^* = 143. \end{aligned}$$