University of California, Los Angeles
Department of Statistics

Statistics 13 **Instructor: Nicolas Christou**

Exam 2
19 February 2016

Name: SOLUTIONS **UCLA ID:** ____________ **Section:** ______

Problem 1 (25 points)
Answer the following questions:

a. Suppose the lifetime of an electronic component follows the exponential distribution with parameter λ. Show that $P(X > s+t|X > t) = P(X > s)$, for $s > 0, t > 0$.

$$P[X > s+t \mid X > t] = \frac{P(X>s+t \cap X>t)}{P(X>t)} = \frac{P(X>s+t)}{P(X>t)}$$

$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = P(X>s).$$

MEMORYLESS PROPERTY OF THE EXPONENTIAL DISTRIBUTION.

b. Suppose we want to estimate the proportion of persons in a large population who have a certain disease. A random sample of 100 persons is selected from the population and the sample proportion of persons in the sample who have the disease is observed. Show that, no matter how large the population is, the standard deviation of the sample proportion is at most 0.05.

$$VAR(\hat{p}) = \frac{p(1-p)}{n} = \frac{0.5(1-0.5)}{100} = 0.0025$$

$$SD(\hat{p}) = \sqrt{0.0025} = 0.05.$$

c. Suppose that evidence concerning the guilt or innocence of a defendant in a criminal investigation can be summarized by the value of an exponential random variable X whose μ depends on whether the defendant is guilty. If innocent, $\mu = 1$; if guilty $\mu = 2$. The deciding judge will rule the defendant guilty if $X > c$ for some suitable chosen value of c. If the judge wants to be 95% certain that an innocent man will not be convicted, what should be the value of c?

THE VALUE OF C SHOULD BE SUCH THAT $P(X>c) = 0.05$

$$e^{-1 \cdot c} = 0.05 \rightarrow \underline{c = 2.996}.$$

d. Refer to question (c). Using the value of c found in question (c), what is the probability that a guilty defendant will be convicted?

$$P(Y>c) = e^{-\frac{1}{2}(2.996)} = 0.2236.$$

$$\lambda = \frac{1}{2}.$$

Problem 2 (25 points)
Answer the following questions:

a. Suppose the number of particles in a sample of water follows the Poisson distribution with $\lambda = 100$ particles per cubic inch. Find the probability that there are at least 125 particles in a volume of 1 cubic inch. Use normal approximation.

$$P(X \geq 125) = P\left(Z > \frac{124.5 - 100}{\sqrt{100}}\right) = P(Z > 2.45) = 0.0071.$$

b. An environmental engineer believes that there are two contaminants in a water supply, arsenic and lead. The actual concentrations of the two contaminants are independent random variables X and Y measured in the same units. The engineer is interested in what proportion of the contaminants is lead on average. That is, the engineer wants to know the mean of $R = \frac{Y}{X+Y}$. A common way to approximate the mean of R is to generate using simulations n pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ and compute $R_i = \frac{Y_i}{X_i+Y_i}$ for $i = 1, 2, \ldots, n$. Then $\bar{R}$ will be a good estimator of the mean of R. Suppose $var(R_i) \approx \frac{1}{4}$. Determine the simulation size needed if we want $\bar{R}$ to be within 0.005 from the true mean of R with probability 98%.

$$P[\mu - 0.005 \leq \bar{R} \leq \mu + 0.005] = 0.98$$

$$P\left[\frac{\mu - 0.005 - \mu}{0.5/\sqrt{n}} \leq Z \leq \frac{\mu + 0.005 - \mu}{0.5/\sqrt{n}}\right] = 0.98$$

$$2.325 = \frac{0.005\sqrt{n}}{0.5} \rightarrow \underline{n = 54056}$$

c. In a large forest the number of pine trees per acre follows the Poisson distribution with $\lambda = 5$. If we randomly select 10 acres from this forest, what is the probability that at least one acre will have at most one pine tree?

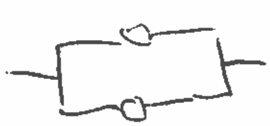
$$P(X \leq 1) = P(X=0) + P(X=1) = \frac{5^0 e^{-5}}{0!} + \frac{5^1 e^{-5}}{1!} = 0.04.$$

$$\left.\begin{array}{l} n = 10 \\ p = 0.04 \end{array}\right\} P(Y \geq 1) = 1 - P(Y=0) = 1 - \binom{10}{0} 0.04^0\, 0.96^{10}$$

$$= 1 - 0.96^{10} = 0.3352.$$

d. The lifetime of a certain electronic component follows the exponential distribution with parameter $\lambda = \frac{1}{100}$. Suppose that 2 such components operate independently and in series in a certain system (that is, the system fails when either component fails). (a). Find the probability that the system operates after 60 hours. (b). Find the probability the system operates after 60 hours if the components are connected in parallel (that is, the system does not fail until both components fail).

$$P(X_1 > 60) \cdot P(X_2 > 60) = e^{-0.6} \cdot e^{-0.6} = 0.3012$$

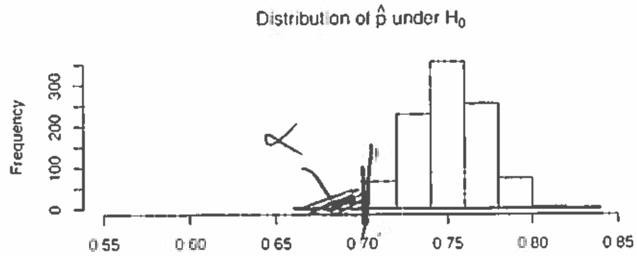


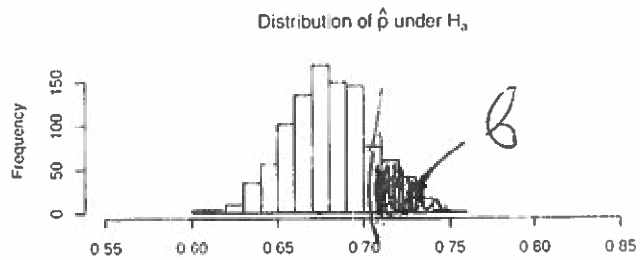
$$1 - P(X_1 < 60) \cdot P(X_2 < 60)$$

$$= 1 - (1 - e^{-0.60})(1 - e^{-0.60}) = 0.7964.$$

Problem 3 (25 points)
Answer the following questions:

a. It was claimed that 75% of all dentists recommend a certain brand of gum for their patients. A consumer group doubted this claim and decided to test $H_0: p = 0.75$ against the alternative hypothesis that $H_a: p < 0.75$ by contacting a survey of 390 dentists. Using the simulation-based approach for inference on the proportion p we constructed the following two histograms (the first one is under the hypothesis that $p = 0.75$ and the second one under the hypothesis that $p = 0.68$). We will reject H_0 if fewer than 273 dentists recommend this brand of gum. Find the Type I error α and the Type II error β and show them on the histograms.





b. Refer to question (a). Write the expressions that compute the exact probability of the Type I error α and the Type II error β.

$$\alpha = P(X < 273) = \sum_{x=0}^{272} \binom{390}{x} 0.75^{x}\, 0.25^{390-x}$$

$$\beta = P(X \geq 273) = \sum_{x=273}^{390} \binom{390}{x} 0.68^{x}\, 0.32^{390-x}$$

c. Refer to question (b). Approximate the probabilities using the normal distribution.

$$\alpha = P\left(Z < \frac{272.5 - 390(0.75)}{\sqrt{390(0.75)(0.25)}}\right) = P(Z < -2.34) = 0.0096$$

$$\beta = P\left(Z > \frac{272.5 - 390(0.68)}{\sqrt{390(0.68)(0.32)}}\right) = P(Z > 0.79) = 0.2148$$

d. For developing countries in Africa and the Americas, let p_1 and p_2 be the respective proportions of babies with low birth weight (below 2500 grams). We will test $H_0: p_1 = p_2$ against the alternative hypothesis $H_a: p_1 > p_2$.

1. Define a critical region that has $\alpha = 0.05$. $Z > 1.645$
2. If random samples of sizes $n_1 = 900$ and $n_2 = 700$ gave $x_1 = 135$ and $x_2 = 78$ babies with low birth weight, what is your conclusion?

$$\hat{p}_1 = \frac{135}{900} = 0.15$$

$$\hat{p}_2 = \frac{78}{700} = 0.11$$

$$\hat{p} = \frac{135 + 78}{900 + 700} = 0.13$$

$$z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{900} + \frac{1}{700}\right)}} = 2.36$$

REJECT H_0 BECAUSE $2.36 > 1.645$

Problem 4 (25 points)
Answer the following questions:

a. Consider the regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, with $\epsilon_i \sim N(0, \sigma)$, and $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ independent. Suppose β_0 is known, and therefore there is only one parameter to estimate, β_1. It can be shown that using the method of least squares, $\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - \beta_0 \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2}$. Find $E(\hat{\beta}_1)$ and $var(\hat{\beta}_1)$. What is the distribution of $\hat{\beta}_1$?

$$E\hat{\theta}_1 = \frac{\Sigma x_i(\theta_0 + \theta_1 x_i) - \theta_0 \Sigma x_i}{\Sigma x_i^2} = \theta_1 \frac{\Sigma x_i^2}{\Sigma x_i^2} = \theta_1.$$

$$var(\hat{\theta}_1) = \frac{\sigma^2 \Sigma x_i^2}{(\Sigma x_i^2)^2} = \frac{\sigma^2}{\Sigma x_i^2}$$

$$\hat{\theta}_1 \sim N\left[\theta_1, \frac{\sigma}{\sqrt{\Sigma x_i^2}}\right]$$

b. A health insurance policy costs \$1000, and will pay policyholders \$5000 if they suffer an injury (resulting in hospitalization). The company estimates that each year 1 in every 10 policyholders may have this kind of injury. Use the central limit theorem to compute the probability that the insurance company will make more than \$28000 if they sell 64 such policies.

x	$p(x)$
1000	0.90
-4000	0.1

$EX = 500$
$var(x) = 2250000$
$SD(x) = 1500$

$$P(T > 28000) = P\left(Z > \frac{28000 - 64(500)}{1500\sqrt{64}}\right)$$
$$= P(Z > -0.33) = 1 - 0.3707$$
$$= 0.6293$$

c. Refer to question (b). Find the expression that computes the exact probability that the insurance company will make more than \$28000 if they sell 64 such policies. Approximate this probability using the normal distribution.

$64 \;\begin{cases} K \\ 64-K \end{cases}$

$$-4000(64-K) + 1000K > 28000 \rightarrow K \geq 57$$

$$P(Y \geq 57) = \sum_{y=57}^{64} \binom{64}{y} 0.9^y \, 0.1^{64-y}.$$

$$P\left(Z > \frac{56.5 - 64(0.90)}{\sqrt{64(0.9)(0.1)}}\right) = P(Z > -0.46) = 1 - 0.3228 = 0.6772.$$

d. Suppose that $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are two independent random samples, with $X_i \sim N(\mu_1, \sigma_1)$ and $Y_i \sim N(\mu_2, \sigma_2)$. Suppose $\sigma_1^2 = 2, \sigma_2^2 = 2.5$. Find the sample size n so that $\bar{X} - \bar{Y}$ will be within one unit of $\mu_1 - \mu_2$ with probability 0.95.

$$P(\mu_1 - \mu_2 - 1 \leq \bar{X} - \bar{Y} \leq \mu_1 - \mu_2 + 1) = 0.95$$

$$P\left(\frac{\mu_1 - \mu_2 - 1 - (\mu_1 - \mu_2)}{\sqrt{\frac{2}{n} + \frac{2.5}{n}}} \leq Z \leq \frac{\mu_1 - \mu_2 + 1 - (\mu_1 - \mu_2)}{\sqrt{\frac{2}{n} + \frac{2.5}{n}}}\right) = 0.95$$

$$1.96 = \frac{1}{\sqrt{\frac{4.5}{n}}}$$

$$n = 17.3 \simeq 18.$$