PHP 2510 (Fall, 2009)

Second Midterm Exam

November 17

Name: ______________________________

Solutions

Note: Write your name on the top of each page. There are three questions on this exam. Show your reasoning for each answer, and circle your final answer to avoid any ambiguities. Writing just the answer without explanation will not receive full credit, even if the answer is correct.

1. The buyer of roots that contain insecticide wants assurance that the average content of the active ingredient is at least 8%, apart from a 5-in-100 chance. A sample of 36 bundles of roots gives, on analysis, $\bar{X} = 8.9\%$. If $\sigma = 3\%$, answer the following questions.

   (a) Calculate the standard error of sample mean.

   (b) Suppose that the central limit theorem is applicable here. Let μ denote the true percentage of the active ingredient. Write down the distribution of sample mean $\bar{X}$ implied by the theorem?

   (c) Calculate a two-sided 95% confidence interval for μ. Does the batch meet the specification?

   (d) We can also use a one-sided 95% confidence interval to answer the question. Which side should we use; calculate the appropriate one? Does the batch meet the specification?

   (e) Explain what the 95% coverage probability means in the context using no more than five sentences.

   (f) If in another assessment, we want the width of a two-sided 95% CI to be 0.98, how many bundles should we include in the sample?

(a). standard error $= 3/\sqrt{36} = 0.5\%$

(b). $\bar{X} \sim N(\mu, 9/36)$

(c): A 95% CI =

$$\bar{x} \pm z_{\alpha/2} \cdot \text{std.err}$$
$$= 8.9 \pm 1.96 \times 0.5$$
$$= 8.9 \pm 0.98$$
$$= (7.92, \ 9.88)$$

Conclusion: 8% is inside the 95% CI. The batch does not meet the specification.

Name:

(d). A one-sided 95% CI should be

$$(\bar{X} - z_\alpha \cdot \text{std.err}, \infty)$$
$$= (8.9 - 1.64 \times 0.5, \infty)$$
$$= (8.9 - 0.82, \infty)$$
$$= (8.08, \infty)$$

Now, 8% is outside the 95% CI. This suggests that the batch meets the specification.

(Using (c) and (d), we obtain different conclusion. People use both to this type of testing. Strictly speaking, we should use (d). (c) is more conservative.

(e). If we repeat the experiment many many times and calculate a 95% CI for each experiment, only 95% of these CI's contain μ.

(f) The width is

$$2 \times 1.96 \times \frac{3}{\sqrt{n}} = 0.98$$

$$\Rightarrow \quad 2 \times 2 \times 3 = \sqrt{n}$$

$$\Rightarrow \quad n = 144.$$

2. Assume that the number of episodes per year of otitis media (a common disease of the middle ear in early childhood) follows a Poisson distribution with parameter λ, i.e.

$$\Pr(X) = \frac{e^{-\lambda}(\lambda)^X}{X!},$$

where X denotes the number of episodes each year and $\lambda = EX$ is the average number of episodes per year.

(a) What is the probability of seeing 2 episodes if $\lambda = 2$.

For the following questions, suppose that λ is unknown and will be estimated by maximum likelihood. Let X_1, X_2, X_3, X_4, and X_5 denote the numbers of episodes from 2001-2005. Suppose these counts are independent.

(b) What is the joint probability of $X_1, X_2, ..., X_5$? What is the likelihood function $L(\lambda \mid X_1, ..., X_5)$? What is the different between them?

(c) Find the score function $U(\lambda)$, and show that it is only a function of T where $T = \sum_{i=1}^{5} X_i$

(d) For (d) and (e), suppose that $T = 8$. Find the MLE $\hat{\lambda}_{\text{MLE}}$ by solving $U(\lambda) = 0$.

(e) Show that the information is

$$I(\lambda) = \frac{5}{\lambda},$$

and find estimate of $var(\hat{\lambda}_{\text{MLE}})$.

(a) $$Pr(x=2) = \frac{e^{-2}\, 2^2}{2!} = 2 \cdot e^{-2} = 0.27$$

(b). $$Pr(x_1, x_2, \cdots, x_5) = \prod_{i=1}^{5} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$= \frac{1}{\prod_{i=1}^{5} x_i!}\, e^{-5\lambda} \cdot \lambda^{\sum_{i=1}^{n} x_i}$$

$$L(\lambda \mid x_1, x_2, \cdots x_5) = \prod_{i=1}^{5} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$= \frac{1}{\prod_{i=1}^{5} x_i!}\, e^{-5\lambda} \cdot \lambda^{\sum_{i=1}^{n} x_i}$$

They take the same form, but $Pr(x_1, \cdots, x_5)$ is a function of $x_1, x_2, \cdots, x_5$. while $L(\lambda \mid x_1, x_2, \cdots, x_5)$ is a function of λ.

Name:

(c) $\log L(\lambda | x_1, x_2, \cdots, x_5) = -\sum_{i=1}^{5} \log x_i! - 5\lambda + \sum_{i=1}^{5} x_i \log\lambda$

$$\Rightarrow U(\lambda | x_1, x_2, \cdots, x_5) = \frac{d \log L}{d\lambda}$$

$$= -5 + \frac{\sum_{i=1}^{5} x_i}{\lambda}$$

$$= -5 + \frac{T}{\lambda}$$

(d) Set $U(\lambda) = 0$

$$\Rightarrow -5 + \frac{T}{\lambda} = 0 \quad \Rightarrow \quad \hat{\lambda}_{MLE} = \frac{T}{5}$$

So, $\hat{\lambda}_{MLE} = 8/5 = 1.6$

(e). $$I(\lambda) = -E\left\{\frac{d^2 \log L(\lambda)}{d\lambda^2}\right\}$$

$$= -E\left\{\frac{d\, U(\lambda)}{d\lambda}\right\}$$

$$= -E\left\{-\frac{T}{\lambda^2}\right\}$$

Here $ET = \sum_{i=1}^{5} E(X_i) = \sum_{i=1}^{5} \lambda = 5\lambda$

$$\Rightarrow I(\lambda) = \frac{5\lambda}{\lambda^2} = \frac{5}{\lambda}$$

$\Rightarrow Var(\hat{\lambda}_{MLE}) = I^{-1}(\lambda) = \frac{\lambda}{5}$. We substitute the unknown λ by $\hat{\lambda}_{MLE}$, and estimate $Var(\hat{\lambda}_{MLE}) = \frac{1.6}{5} = 0.32$.

3. Iron-deficiency anemia is an important nutritional health problem in the US. A dietary assessment was performed on 36 boys 9-11 years old whose families were below the poverty level. The mean daily iron intake among these boys was found to be 12.5 mg with sample standard deviation 4.8 mg. Suppose the mean daily iron intake among the population of 9- to 11-year-old boys from all income strata is 14 mg. We want to test if the mean iron intake among the low-income group is lower than that of the general population.

   (a) State the hypotheses that we can use to consider this question. Should you use one-sided or two-sided test?

   (b) Carry out the hypothesis test in (a) using the critical-value method with an α level of 0.05. Summarize your findings.

   (c) Explain what $\alpha = 0.05$ means?

   (d) What is the p-value for the test conducted in (b).

   (e) If the mean iron intake among the low-income group is actually 10 mg, what is the power of carrying out the study with a one-sided α level of 0.05?

   (f) If we double the sample size to 72, will the power increase or decrease (no calculation is needed)? Briefly explain why (no more than 5 sentences; you can draw a picture(s) to help you explain).

(a). Let μ denote the mean daily iron intake of children from poor families.

$H_0: \mu = 14$ vs $H_1: \mu < 14$.

We should use a one-sided test.

(b). Assume H_0 is true.

$$c = 14 - z_\alpha \sigma/\sqrt{n} = 14 - 1.64 \times 4.8/\sqrt{36} = 12.7$$

$\bar{x} = 12.5 < c$, so we reject the H_0 and favor the H_1 at a significance level < 0.05.

(c) The probability of mistakenly rejecting the H_0 when the H_0 is actually true is 0.05.

(d) When the H_0 is true,

$$z = \frac{\bar{x} - 14}{4.8/\sqrt{36}} \sim N(0,1)$$

$$z = \frac{12.5 - 14}{0.8} = -1.5/0.8 = -1.875$$

The p-value for a one-side test is 0.03.

(e). From (c), we will reject the H_0 if $\bar{x} < c$. When $\mu = 10$, the power is

$$\text{power} = Pr(\bar{x} < c \mid H_1 \text{ is true}).$$

$$= Pr\left(\frac{\bar{x} - 10}{4.8/\sqrt{n}} < \frac{c - 10}{4.8/\sqrt{n}} \,\middle|\, H_1\right)$$

$$= Pr\left(z < \frac{12.7 - 10}{0.8}\right), \text{ when } z \sim N(0,1)$$

$$= Pr(z < 3.375) = 0.999$$

(f). From (e). we know that

$$\text{power} = Pr\left(\frac{\bar{x} - 10}{4.8/\sqrt{n}} < \frac{c - 10}{4.8} \cdot \sqrt{n} \,\middle|\, H_1\right)$$

$$= Pr\left(z < \frac{c - 10}{4.8} \cdot \sqrt{n}\right)$$

where $c = 14 - 1.64 \times \frac{4.8}{\sqrt{n}}$.

As n increases, c increases, and $\frac{c - 10}{4.8}$ also increases. So, $\frac{c - 10}{4.8}\sqrt{n}$ becomes a larger number. Hence, the power increases.