STA 261s2005 Assignment 8

Do this assignment in preparation for the quiz on Wednesday, March 16th. The questions are practice for the quiz, and are not to be handed in.

Please start by reading Sections 12.1 and 12.2 in the text, and also the online supplement, pages 5-11. You are not responsible for the material on uniformly most powerful tests and the Neyman-Pearson Lemma. It will be on the next assignment.

1. Let X_1 and X_2 be independent random variables. Let $Y_1 = g_1(X_1)$ and $Y_2 = g_2(X_2)$. Let the function $g_1(x)$ be strictly increasing, and let the function $g_2(x)$ be strictly decreasing. So, both functions are one-to-one and have unique inverses. Prove that Y_1 and Y_2 are independent.

2. $X_1, \ldots, X_{n_1}$ be a random sample. For each of the distributions below, give the parameter space Θ and the sample space $\mathcal{X}$

   (a) Bernoulli

   (b) Binomial(k, θ) with k known

   (c) Binomial(k, θ) with k unknown

   (d) Poisson

   (e) Geometric

   (f) Uniform(α, β)

   (g) Exponential

   (h) Gamma

   (i) Normal

3. Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with expected value μ and variance σ^2.

   (a) We will test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ using the critical region

   $$C = \{\mathbf{x} \in \mathcal{X} : \left| \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \right| > t_{\alpha/2,n-1}\}$$

   This is called a *one-sample t-test.* If the observations are differences (like pretest versus post-test) and $\mu_0 = 0$, it is called a *matched t-test.*

   i. What is Θ_0? Is it simple or composite?

   ii. What is Θ_1? Is it simple or composite?

   iii. What is the size of the test? Prove your answer. Start with the independence of $\overline{X}$ and S^2 and the distribution of $\frac{(n-1)S^2}{\sigma^2}$. You did some of this work in the last assignment; just be ready to do it again.

   iv. Show that H_0 is rejected if and only if the $(1-\alpha)100\%$ confidence interval for μ does *not* include μ_0. It is easiest to start by writing the set of $\mathbf{x} \in \mathcal{X}$ such that μ_0 is in the confidence interval, and then work on it until it becomes $C^c \cap \mathcal{X}$.

v. Suppose the data are not really normal, but the sample size is large. Does the test still have the same approximate size? Answer Yes or No and explain why.

vi. Suppose a large corporation takes a random sample of 200 employees, and then on two *non*randomly selected days, one day with beautiful weather and one day with horrible weather, checks whether each employee called in sick or not. Each employee gets a 1 if (s)he calls in sick, and a zero if (s)he does not. Would it make sense to calculate a *difference* (which must be -1, 0 or 1 and so is definitely not normal) for each employee, and then use a one-sample t-test to decide whether the expected difference has mean zero? Explain.

4. Still for the single sample from a normal distribution, we will test $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$ using the critical region

$$C = \{\mathbf{x} \in \mathcal{X} : \frac{(n-1)S^2}{\sigma_0^2} > \chi^2_{\alpha/2,n-1} \text{ or } \frac{(n-1)S^2}{\sigma_0^2} < \chi^2_{1-\alpha/2,n-1}\}$$

   (a) What is Θ_0? Is it simple or composite?
   (b) What is Θ_1? Is it simple or composite?
   (c) What is the size of the test? Prove your answer.

5. Now we wish to test $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$ using the critical region

$$C = \{\mathbf{x} \in \mathcal{X} : \frac{(n-1)S^2}{\sigma_0^2} > k\}$$

   (a) What is Θ_0? Is it simple or composite?
   (b) What is Θ_1? Is it simple or composite?
   (c) Find k so that the size of the test is α.
   (d) Show that this same test is also size α for testing $H_0 : \sigma^2 \leq \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$.

6. Let $X_1, \ldots, X_{n_1}$ be a random sample from a $N(\mu_1, \sigma^2)$ distribution, and let $Y_1, \ldots, Y_{n_2}$ be a random sample from a $N(\mu_2, \sigma^2)$ distribution. These are *independent* random samples, meaning that the X and Y values are independent. Using the facts that $\overline{X}$ and S^2 are independent and $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ for each sample,

   (a) Give a size α critical region for testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. Hint: Look at the corresponding confidence interval. Your answer is called a *two-sample t-test*, or an *independent t-test*.

      i. What is Θ_0? Is it simple or composite?
      ii. What is Θ_1? Is it simple or composite?

   (b) Give a size α critical region for testing $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$. Hint: Look at the corresponding confidence interval.

      i. What is Θ_0? Is it simple or composite?

ii. What is Θ_1? Is it simple or composite?

7. Let $X_1, \ldots, X_{n_1}$ be a random sample from a (possibly) non-normal distribution with mean μ_1 and variance σ_1^2), and let $Y_1, \ldots, Y_{n_2}$ be a random sample from a (possibly) non-normal distribution with mean μ_2 and variance σ_2^2). These are *independent* random samples, meaning that the data are independent between samples as well as within samples. We are interested in testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. Find the constant k so that the following critical region will have, for large n_1 and large n_2, a size of approximately α.

$$C = \left\{ \mathbf{x} \in \mathcal{X} : \left| \frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\sigma}_2^2}{n_2}}} \right| > k \right\},$$

where $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ are consistent estimators of σ_1^2 and σ_2^2, respectively.

8. Of a random sample of 150 Special Needs students in the Toronto District School Board, 19 were in regular classes, and the rest were in Special Education classes. Of a random sample of 200 Special Needs students in the Toronto Separate School Board, 48 were in regular classes, and the rest were in Special Education classes. Test for difference between the proportions of Special Needs students in regular classes in the two school boards. Use $\alpha = 0.01$. What do you conclude? Of course you should use the test from the last question.

9. Let $X_1, \ldots, X_{n_1}$ be a random sample from an exponential distribution with parameter θ. Give a size α critical region for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Your answer should use chi-square critical values.

10. Let $Y_i = x_i + \epsilon_i$, for $i = 1, \ldots, n$, where

- $x_1, \ldots, x_n$ are fixed, known constants
- $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed Normal(0,σ^2) random variables; the parameter σ^2 is unknown.
- The data consist of n pairs (x_i, Y_i). The error terms ϵ_i are not given directly.

(a) Find the distribution of $\frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - x_i)^2$.

(b) We wish to test $H_0 : \sigma^2 \leq \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$.

i. What is Θ?

ii. What is $\mathcal{X}$?

iii. What is Θ_0? Is it simple or composite?

iv. What is Θ_1? Is it simple or composite?

(c) Find the constant k so that the following test will be size α for testing the *simple* null hypothesis $H_0 : \sigma^2 = \sigma_0^2$.

$$C = \{\mathbf{x} \in \mathcal{X} : \frac{\sum_{i=1}^n (Y_i - x_i)^2}{\sigma_0^2} > k\}$$

(d) Find the power function $P_{\sigma^2}\{\mathbf{X} \in C\} = \pi(\sigma^2)$.

(e) Prove that the test C is also size α for testing $H_0 : \sigma^2 \leq \sigma_0^2$.

11. In this unfamiliar but reasonable regression model, the effect of x is linear as usual, but each member of the population has his or her own individual slope. That makes the slope a random variable, because if you took another sample, you'd get another collection of slopes. Accordingly,

    Let $Y_i = x_i B_i$, for $i = 1, \ldots, n$, where

    - $x_1, \ldots, x_n$ are fixed, known constants
    - $B_1, \ldots, B_n$ are independent and identically distributed Normal(β, σ^2) random variables.
    - The parameters β and σ^2 are unknown.
    - The data consist of n pairs (x_i, Y_i). The slopes B_i are not given directly.
    - Yes, there is no error term. Don't worry.

    We wish to test $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$.

    (a) What is Θ?

    (b) What is $\mathcal{X}$?

    (c) What is Θ_0? Is it simple or composite?

    (d) What is Θ_1? Is it simple or composite?

    (e) Find the distribution of $\frac{1}{n}\sum_{i=1}^n \frac{Y_i}{x_i}$.

    (f) Find the constant k so that the following test will be size α:

$$C = \left\{ \mathbf{x} \in \mathcal{X} : \left| \frac{\sqrt{n(n-1)}[(\frac{1}{n}\sum_{i=1}^n \frac{Y_i}{x_i}) - \beta_0]}{\sqrt{\sum_{i=1}^n [Y_i/x_i - (\frac{1}{n}\sum_{j=1}^n \frac{Y_j}{x_j})]^2}} \right| > k \right\}$$

12. Now modify the last example so that that the slopes have an exponential distribution with parameter $\beta > 0$ (just to clarify, $E[B_i] = \beta$). This is supposed to model a situation where the random slopes are known to be positive. Starting with the distribution of

$$\overline{W} = \frac{1}{n}\sum_{i=1}^n W_i = \frac{1}{n}\sum_{i=1}^n \frac{Y_i}{x_i},$$

    (a) Give a size α critical region for testing $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$. Your answer should use chi-square critical values.

    (b) Prove that your test is also size α for testing $H_0 : \sigma^2 \leq \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$.

13. Let $X_1, \ldots, X_{n_1}$ be a random sample from a uniform distribution on $(0, \theta]$. We wish to test $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$..

    (a) Write down the cumulative distribution function of the sample maximum Y_n; you will need it.

(b) Show that the following test is of size α:

$$C = \left\{ \mathbf{x} \in \mathcal{X} : Y_n > \theta_0 \text{ or } Y_n < \theta_0 \alpha^{1/n} \right\}$$

(c) It makes sense that the critical value $\theta_0 \alpha^{1/n}$ is less than θ_0, but please prove it. This will help later.

(d) Now we will find the power function $P_\theta\{\mathbf{X} \in C\} = \pi(\theta)$

i. Show $\pi(\theta) = 1$ for $\theta < \theta_0 \alpha^{1/n}$.

ii. Show $\pi(\theta) = (\frac{\theta_0}{\theta})^n \alpha$ for $\theta_0 \alpha^{1/n} \leq \theta < \theta_0$.

iii. That looks suspicious because $\frac{\theta_0}{\theta} > 1$. To see that it's okay, show that $(\frac{\theta_0}{\theta})^n \alpha \leq 1$ for $\theta_0 \alpha^{1/n} \leq \theta < \theta_0$.

iv. Show $\pi(\theta) = 1 - (\frac{\theta_0}{\theta})^n (1 - \alpha)$ for $\theta \geq \theta_0$.

(e) Suppose that $\theta_0 = 10$, $\alpha = 0.05$, and the true value of θ is 9. What value of n is required so that the probability of Type II Error is zero? My answer is $n = 29$.

(f) Suppose the true value of θ is *less than* 9 and $n = 29$. Is the probability of Type II Error still zero?

(g) Suppose $\theta > \theta_0$. What sample size is required so that the probability of Type II error is less than β, where $0 < \beta < 1$? My answer is $n > \frac{\ln \theta_0 - \ln \theta_1}{\ln \beta - \ln(1-\alpha)}$.