STA 302f17 Assignment Two[1]

Please bring your R printout from Question 2h to Quiz Two; you may be asked to hand it in, or maybe not. The other problems are preparation for the quiz in tutorial, and are not to be handed in. Starting with Problem 4, you can play a little game. Try not to do the same work twice. Instead, use results of earlier problems whenever possible.

1. This problem is more review, this time of statistical concepts you likely encountered in STA258. Let $y_1, \ldots, y_n$ be a random sample[2] from a normal distribution with mean μ and variance σ^2, so that $T = \frac{\sqrt{n}(\overline{y}-\mu)}{S} \sim t(n-1)$. This is something you don't need to prove, for now.

   (a) Derive a $(1-\alpha)100\%$ confidence interval for μ. "Derive" means show all the high school algebra. Use the symbol $t_{\alpha/2}$ for the number satisfying $Pr(T > t_{\alpha/2}) = \alpha/2$.

   (b) A random sample with $n = 23$ yields $\overline{y} = 2.57$ and a sample variance of $S^2 = 5.85$. Using the critical value $t_{0.025} = 2.07$, give a 95% confidence interval for μ. The answer is a pair of numbers, the lower confidence limit and the upper confidence limit. Please **bring a calculator to the quiz** in case you have to do something like this.

   (c) Using the numbers from Question 1b, test $H_0 : \mu = 3$ at $\alpha = 0.05$.

      i. Give the value of the T statistic. The answer is a number.

      ii. State whether you reject H_0, Yes or No.

      iii. Can you conclude that μ is different from 3? Answer Yes or No.

      iv. If the answer is Yes, state whether $\mu > 3$ or $\mu < 3$. Pick one.

2. In the textbook *Regression Analysis*, please read pages 1-7 for general ideas. Then read Section 1.4 on pages 7-9. The chapter has an appendix with some derivations, too.

   (a) In formula (1.5), assume $E(\epsilon_i) = 0$. What is $E(y_i)$?

   (b) Partially differentiate expression (1.8) to obtain formulas (1.11) and (1.13) for b_0 and b_1 on page 8.

   (c) Prove that (1.13) and (1.14) are equal.

   (d) On p. 9, the text says $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$. Prove it.

   (e) For the centered model (1.15), partially differentiate to obtain the the least squares estimates of γ_0 and β_1. Is the least-squares estimate of the slope affected by centering?

[1]Copyright information is at the end of the last page.

[2]Random sample means independent and identically distributed.

(f) For the simple regression model (1.5), show that the residuals add to zero as claimed on p. 9..

(g) Do Exercise 1.6. See the definition of least squares estimation in the lecture slides.

(h) Do Exercise 1.9 using R. I got the data in using the `c()` function – `c` for collect. The problem is asking you to calculate b_0 and b_1. **Bring your printout to the quiz. You may be asked to hand it in.**

Note that while the textbook gives $\log(y)$ rounded to two decimal places, you don't need to round if you are using R – so please don't round. When the text says $\log(y)$, does this mean the natural log, or log base ten? One more comment is that if you plot x versus y (not requested by this question), you see a clearly curvy relationship, while a plot of x versus $\log(y)$ is very close to a straight line. Transformation of y is a good curve-fitting trick.

3. Please read Section 1.5 in the textbook *Regression Analysis*. Assume $E(\epsilon_i) = 0$.

(a) In formula (1.20), what is $E(y_i)$?

(b) Obtain the least squares estimator of β_1 for this model. Show all your work including the second derivative test.

(c) Do Exercise 1.1.

(d) What is $\sum_{i=1}^n e_i$ for this model? Must it be equal to zero? Answer Yes or No.

(e) What is $E(e_i)$? What is $E(\sum_{i=1}^n e_i)$?

4. Denote the moment-generating function of a random variable y by $M_y(t)$. The moment-generating function is defined by $M_y(t) = E(e^{yt})$.

(a) Let a be a constant. Prove that $M_{ax}(t) = M_x(at)$.

(b) Prove that $M_{x+a}(t) = e^{at}M_x(t)$.

(c) Let $x_1, \ldots, x_n$ be *independent* random variables. Prove that

$$M_{\sum_{i=1}^n x_i}(t) = \prod_{i=1}^n M_{x_i}(t).$$

For convenience, you may assume that $x_1, \ldots, x_n$ are all continuous, so you will integrate. This is not the way I did it in class.

5. Recall that if $x \sim N(\mu, \sigma^2)$, it has moment-generating function $M_x(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$.

(a) Let $x \sim N(\mu, \sigma^2)$ and $y = ax + b$, where a and b are constants. Use moment-generating functions to find the distribution of y. Show your work.

(b) Let $x \sim N(\mu, \sigma^2)$ and $z = \frac{x-\mu}{\sigma}$. Find the distribution of z. Show your work.

(c) Let $x_1, \ldots, x_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. Use moment-generating functions to find the distribution of $y = \sum_{i=1}^n x_i$. Show your work.

(d) Let $x_1, \ldots, x_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. Use moment-generating functions to find the distribution of the sample mean $\overline{x}$. Show your work.

(e) Let $x_1, \ldots, x_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. Find the distribution of $z = \frac{\sqrt{n}(\overline{x}-\mu)}{\sigma}$. Show your work.

(f) Let $x_1, \ldots, x_n$ be independent random variables, with $x_i \sim N(\mu_i, \sigma_i^2)$. Let $a_1, \ldots, a_n$ be constants. Use moment-generating functions to find the distribution of $y = \sum_{i=1}^n a_i x_i$. Show your work.

6. For the model of formula (1.20) in the text, suppose that the ϵ_i are normally distributed and independent, which is the usual assumption. What is the distribution of y_i? What is the distribution of b_1? Use earlier work to obtain the answers without directly using moment-generating functions.

7. A Chi-squared random variable x with parameter $\nu > 0$ has moment-generating function $M_x(t) = (1-2t)^{-\nu/2}$.

(a) Let $x_1, \ldots, x_n$ be independent random variables with $x_i \sim \chi^2(\nu_i)$ for $i = 1, \ldots, n$. Find the distribution of $y = \sum_{i=1}^n x_i$.

(b) Let $z \sim N(0,1)$. Find the distribution of $y = z^2$. For this one, you need to integrate. Recall that the density of a normal random variable is $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. You will still use moment-generating functions.

(c) Let $x_1, \ldots, x_n$ be random sample from a $N(\mu, \sigma^2)$ distribution. Find the distribution of $y = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$.

(d) Let $y = x_1 + x_2$, where x_1 and x_2 are independent, $x_2 \sim \chi^2(\nu_2)$ and $y \sim \chi^2(\nu_1 + \nu_2)$, where ν_1 and ν_2 are both positive. Show $x_1 \sim \chi^2(\nu_1)$.

(e) Let $x_1, \ldots, x_n$ be random sample from a $N(\mu, \sigma^2)$ distribution. Show

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1),$$

where $s^2 = \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{n-1}$. Hint: $\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \overline{x} + \overline{x} - \mu)^2 = \ldots$ For this question, you may use the independence of $\overline{x}$ and s^2 without proof. We will prove it later. Note: This is a special case of a central result that will be used throughout most of the course.
