STA 302f15 Assignment Two[1]

These problems are preparation for the quiz in tutorial, and are not to be handed in. Starting with Problem 5, you can play a little game. Try not to do the same work twice. Instead, use results of earlier problems whenever possible.

1. In *Linear models in statistics*, do problems 2.6 (a,c,d), 2.7, 2.17, 2.18, 2.20, 2.23, 2.24 and 2.25. The answers in the back of the book are helpful. The answer to 2.24 is incomplete; consider all four cases.

2. Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ $\mathbf{B} = \begin{pmatrix} 0 & 2 \\ 2 & 1 \end{pmatrix}$ $\mathbf{C} = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$

   (a) Calculate $\mathbf{AB}$ and $\mathbf{AC}$

   (b) Do we have $\mathbf{AB} = \mathbf{AC}$?

   (c) Prove $\mathbf{B} = \mathbf{C}$. Show your work.

   The idea for this problem comes from Example 2.4b (page 21) in the textbook. Also see Problem 2.25 in the text.

3. Sometimes, you want the least squares line to go through the origin, so that predicted Y automatically equals zero when $x = 0$. For example, suppose the cases are n half-kilogram batches of rice purchased from grocery stores. The independent variable x is concentration of arsenic in the rice before washing, and the dependent variable Y is concentration of arsenic after washing. Discounting the very unlikely possibility that arsenic contamination can happen *during* washing, you want to use your knowledge that zero arsenic before washing implies zero arsenic after washing. You will use your knowledge by building it into the statistical model.

   Accordingly, let $Y_i = \beta x_i + \epsilon_i$ for $i = 1, \ldots, n$, where $\epsilon_1, \ldots, \epsilon_n$ are a random sample (that is, independent and identically distributed) from a distribution with expected value zero and variance σ^2, and β and σ^2 are unknown constants. The numbers $x_1, \ldots, x_n$ are known, observed constants.

   (a) What is $E(Y_i)$?

   (b) What is $Var(Y_i)$?

   (c) Find the Least Squares estimate of β by minimizing the function

   $$Q(\beta) = \sum_{i=1}^{n} (Y_i - \beta x_i)^2$$

   over all values of β. Let $\widehat{\beta}$ denote the point at which $Q(\beta)$ is minimal.

   (d) Give the equation of the least-squares line. Of course it's the *constrained* least-squares line, passing through $(0, 0)$.

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

(e) Calculate $\widehat{\beta}$ for the following data set. Your answer is a number. Bring a calculator to the quiz in case you have to do something like this.

```
x  0.0  1.3  3.2 -2.5 -4.6 -1.6  4.5  3.8
y -0.8 -1.3  7.4 -5.2 -6.5 -4.9  9.9  7.2
```

(f) Recall that a statistic is an *unbiased estimator* of a parameter if the expected value of the statistic is equal to the parameter. Is $\widehat{\beta}$ an unbiased estimator of β? Answer Yes or No and show your work.

(g) What is $Var(\widehat{\beta})$? Show your work.

4. Let $Y_1, \ldots, Y_n$ be a random sample 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.

   (c) 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.

5. 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.

6. 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. 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. 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. 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. Find the distribution of $Y = \sum_{i=1}^n a_i X_i$. Show your work.

7. For the model of Question 3, suppose that the ϵ_i are normally distributed, which is the usual assumption. What is the distribution of Y_i? What is the distribution of $\widehat{\beta}$? You should be able to just write down the answers based on your earlier work.

8. 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, $X_1 \sim \chi^2(\nu_1)$ and $Y \sim \chi^2(\nu_1 + \nu_2)$, where ν_1 and ν_2 are both positive. Show $X_2 \sim \chi^2(\nu_2)$.

(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: $\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n \left(X_i - \overline{X} + \overline{X} - \mu\right)^2 = \ldots$ For this question, you may use the independence of $\overline{X}$ and S^2 without proof. We will prove it later.