BUEC 333-D200, Test 1

June 15, 2016, 14:30-17:20

1 Week 1: Probability theory

Consider the joint probability distribution of the random variables X_1 and X_2 given in Table 1.

	$X_2 = 1$	$X_2 = 2$	$X_2 = 3$
$X_1 = 0$	0.2	0.1	0.1
$X_1 = 1$	0.3	0.3	0

Table 1: Joint probability distribution of X_1 and X_2.

1. What is the probability that $X_1 = 0$ given that $X_2 = 2$? **P(X1=0|X2=2)=P(X1=0,X2=2)/P(X2=2)=0.1/(0.1+0.3**
2. What is the conditional expectation of X_1 given that $X_2 = 2$? **P(X1=1|X2=2)=1-P(X1=0|X2=2)=0.75. So E(X1|X2=2)=0.75**
3. Compute $\text{Var}\left(X_1 \middle| X_2 = 3 \right)$. **Since P(X1=0|X2=3)=1, the conditional variance is 0.**
4. Compute the covariance between X_1 and X_2. Get as far as you can without a calculator. **E(X1)=0.6. E(X2)=0.5+2*0.4+3*0.1=1.6. Cov(X1,X2)=0.2*(1-1.6) ...**

2 Week 2: Sampling

Suppose that $Y_1, \cdots, Y_n$ are independent random variables with a common mean μ_Y, but with different variances:

$$\begin{aligned} \text{Var}(Y_1) &= \sigma_1^2 \\ \text{Var}(Y_2) &= \sigma_2^2 \\ &\vdots \\ \text{Var}(Y_n) &= \sigma_n^2 \end{aligned}$$

1. Are the random variables $(Y_1, \cdots, Y_n)$ i.i.d.? **No, they are not "identical", as they have different variances.**
2. Suppose that $n = 3$. Show that $E(\bar{Y}) = \mu_Y$. **E((Y1+Y2+Y3)/3)=1/3*(E(Y1)+E(Y2)+E(Y3)=1/3*3*mu=mu.**
3. Suppose that $n = 2$. What is the variance of the sample mean $\bar{Y}$? **Var(1/2*Y1 + 1/2*Y2)=1/4 sigma^2_1+1/4 sigma^2_2**
4. For qeustions 4 and 5, assume that $n = 2$ and that $\sigma_1^2 = 1$ and that $\sigma_2^2 = 100$.
   Consider the estimator $\tilde{Y} = Y_1$. Is $\tilde{Y}$ unbiased? **Yes, since E(Y-tilde)=E(Y1)=mu_Y**
5. Show that $\tilde{Y}$ is more efficient than $\bar{Y}$. **Var(Y-tilde)=Var(Y1)=sigma^2_1=1. Using (3), Var(Y-bar)=100/4+1/4=25.25>1. Since Var(Y-tilde)<Var(Y-bar), T-tilde is more efficient.**

3 Week 3: Statistics

For this question, remember that a Bernoulli random variable Y has $\text{Var}(Y) = p(1-p)$, where $p = P(Y = 1)$. Therefore, given an estimator $\hat{p}$ for p, the estimator for the variance is $s_Y = \hat{p}(1 - \hat{p})$.

"Would you use the library more if the hours were extended?" In a random sample of 100 SFU freshmen, a sample proportion of $\hat{p}_1 = 0.5$ answered that they would use the library more. A random sample of 100 SFU sophomores was also obtained. Out of them, 80 responded that they would use the library more, for a sample proportion of $\hat{p}_2 = 0.8$.

1. Denote by p_1 the population proportion of SFU freshmen that would use the library more if hours were extended. Construct a 95% confidence interval for p_1. (As far as you can get without a calculator.) **Lower bound of 95%-CI is given by X-bar - 1.96 SE(X) which in this case is p-hat - 1.96*phat(1-phat)/sqrt(n) = 0.5-1.96*0.5*(0.5) / 10. Similar for upper bound.**

2. Test the null hypothesis $H_0 : p_1 = 0.10$ against the two-sided alternative $H_1 : p_1 \neq 0.10$. (Answer this question without computing the $p-$value: you do not have access to the book's Table 1!) **From your computations under (1), you can see that 0.1 is not in the confidence interval. By the relationship between CIs and hyp-testing, you would reject the hypothesis in the current question.**

3. What is the probability that p_1, the true population proportion, is in the interval computed in (1)? If you did not answer the previous question, assume that the correct answer is $[0.4, 0.6]$. **It is either in there or not. Incorrect answer: 95%.**

4. How would you test whether there is a statistically significant difference between freshmen and sophomores in their response? **Say something about section 3.4, including the standard error of the differences, t-ratio, etc.**

4 Week 4: Linear regression (mechanics)

You are given the following random sample:

i	X_1	Y_i
1	0	8
2	2	4
3	1	6

Table 2: Sample for week 4 question.

For this sample:

1. Compute $\hat{\beta}_1$. **-2**

2. Compute the sum of squared residuals. **They are all 0**

3. Without changing the values of X_1, X_2, X_3 and Y_1, give values for Y_2, Y_3 that give $R^2 = 0$. **Y2=Y3=8 would yield betahat1=0 which implies R2=0.**

5 Week 5: Linear regression (statistics)

The questions is about the results of a regression of the weight of infants (newborn babies) and the smoking behavior of the mother. We have data on $n = 1388$ pregnancies. The dependent variable is *bwght*, the weight of the baby when it is born, in ounces. The explanatory variable is *cigs*, the number of cigarettes smoked per day by the mother. We obtain the following results:

$$\widehat{bwght} = \underset{(20.6)}{119.77} - \underset{(0.1)}{0.5} \times cigs\,.$$

Answer the following questions:

1. The underlying model is $bwght_i = \beta_0 + \beta_1 \times cigs_i + u_i$. Give two examples of factors that are captured by u_i. Use factors that are specific to this example, so "it is an error term" or "other stuff" are not correct answers. **Other lifestyle choices, gestation time, ...**

2. State Least Squares Assumption 1 using either math or words. **The factors mentioned under 1 are not related to the number of cigarettes smoked by the mother. Or E(ui|Xi)=0.**

3. Construct a 95% confidence interval for β_1. **[-0.5-1.96*0.1,-0.5+1.96*0.1]**

4. What is the interpretation of β_1, assuming that Assumption 1 holds? **For every additional cigarette smoked, the conditional expectation of birthweight down by 0.5 ounces, ceteris paribus.**

5. What is the interpretation of β_1, assuming that Assumption 1 does not hold? **We can use beta1 to predict a baby's birthweight if we know how many cigarettes the mother smokes**

6 Self-study

1. [Section 4.5] Under the Least Squares Assumptions, what is the expectation of $\hat{\beta}_1$? Is $\hat{\beta}_1$ consistent? **See book.**

2. What is a cumulative probability distribution? Use math and words. **See book.**