Stat 400 : Discussion section BD3 and BD4 Handout 10

Subhadeep Paul

April 23, 2013

1 Conceptual mistakes in midterm

- ***Interpretation of confidence interval :*** see handout 8 for a detailed discussion. CI says nothing about whether the true parameter is contained in a particular interval (or the probability of that), it just comment on how good the method of obtaining the CI is i.e how much are we *confident* about our interval. True parameter value is a constant and not a random variable so we can't talk about assigning probabilities to it. Interpretation of Confidence interval is that if ***we repeat the experiment 100 times, draw samples and construct 95% confidence intervals, 95 times we will come up with an interval that will contain the true value of the parameter***.

- ***p value of two sided tests:*** In simple terms, p-value is the probability value from z or t table corresponding to the test statistic. Think about p-value as similar to α. You reject a null hypothesis if z score is greater than z_α. Similarly you reject the test statistic if p value is smaller than α. Now for two tailed tests, you replace α by $\alpha/2$, to take care of both the possibilities that the test statistic can be sufficiently large either in the right tail or in the left tail of Normal pdf. So while calculating p-value, the one you would get by looking into the table is the one corresponding to $\alpha/2$. That's why you need to multiply it by 2.

- ***minimum sample size in proportions :*** Keep in mind we want to be conservative and take more than enough samples. The term "minimum" is kind of a formality, because if that is not mentioned then one can take infinite sample. But we really want to be enough conservative. Hence if no information about $p*$ is supplied, we will take it as 0.5 as that makes n largest. If an interval is mentioned as possible values of $p*$ we will take the value of $p*$ that maximizes n.

- ***conditions of independence:*** Either $f(x,y) = f_X(x)f_Y(y)$ both (1) The support is a rectangle and (2) $f(x,y)$ can be written as a product of two functions, one of which doesn't contain x and the other one doesn't contain y.

- ***MLE*** The product is over x_1, x_2,....x_n . and not x. so you write Πx_i, and not x^n. The reason being they are random sample, not identical copies and you are trying to find their joint pdf/pmf.

- ***Double integrals*** In those problems, if you can't find the bounds through algebra, draw a figure. Keep in mind all you need to do is to integrate the pdf/pmf over the required area. Example in class.

2 2 Proportions : CI

- ***Interpretation(p):*** Rejecting the null hypothesis is equivalent to saying (when null is $p_1 = p_2$) , $p_1$is significantly different (or greater or smaller) from p_2, which in turn in equivalent to saying $p_1 - p_2$ is (significantly) different from 0. This statement is same as saying 0 is not contained in the confidence interval of $p_1 - p_2$.

Problem 1. (midterm 3, fall 12) A Pentagon statistician is evaluating a prototype bomber to see if it can strike on target more often than the existing bomber can. Two independent samples of size 70 and 75 are obtained

Table 1: Two proportions

Confidence Interval		Hypothesis Testing	
Point estimate	$\hat{p_1} = \frac{x_1}{n_1}, \hat{p_2} = \frac{x_2}{n_2}$	Point estimate	$\hat{p_1} = \frac{x_1}{n_1}, \hat{p_2} = \frac{x_2}{n_2}$
standard error(SE)	$\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}$	standard error(SE)	$\sqrt{\hat{p*}(1-\hat{p*})(1/n_1 + 1/n_2)}$
distribution	Normal	$p*$	$\frac{x_1+x_2}{n_1+n_2}$
total error term ϵ	$z_{\alpha/2}SE$	distribution	Normal
CI	$\hat{p_1} - \hat{p_2} \pm z_{\alpha/2}SE$	Test satistic	$\frac{\hat{p_1} - \hat{P_2}}{SE}$

Table 2: Rejection regions Two proportions and means

Null (p)	Alternative(p)	Null (μ)	Alternative(μ)	Test	Rejection region	p-value
$p_1 = p_2$, $p_1 \leq p_2$	$p_1 > p_2$	$\mu_1 = \mu_0$, $\mu_1 \leq \mu_0$	$\mu_1 > \mu_0$	Two tailed	$z > z_\alpha$	$P(Z \geq z)$
$p_1 = p_2$, $p_1 \geq p_2$	$p_1 < p_2$	$\mu_1 = \mu_0$, $\mu_1 \geq \mu_0$	$\mu_1 < \mu_0$	Left tailed	$z < -z_\alpha$	$P(Z \leq z)$
$p_1 = p_2$	$p_1 \neq p_2$	$\mu_1 = \mu_0$	$\mu_1 \neq \mu_0$	Right tailed	$\lvert z \rvert > z_{\alpha/2}$	$2 * P(Z \geq z)$

with number of successful hits being 41 and 56 respetively. The military officials want to be 95% confident that the prototype is superior to the existing bomber. (a) State the null and alternative hypothesis statements. (b). Compute the appropriate test statistic. (c) Is there evidence to reject the null hypothesis? State your conclusion and supporting evidence.

3 Hypothesis testing for μ and σ

Problem 2. In the past, the mean monthly long-distance telephone bill in a certain area was \$17.85. After an advertising campaign, a random sample of 20 household bills was taken, the sample mean was \$19.35 and sample standard deviation of \$3.88. Assume that the phone bills are approximately normally distributed. (a) Is there enough evidence that the mean monthly long-distance telephone bill has changed. Perform the appropriate test at a 1% significance level. Report p value. (b) Construct a 95% confidence interval for the overall mean monthly long-distance telephone bill. (c) Construct a 95% confidence interval for the overall standard deviation of monthly long-distance telephone bills.

Table 3: Hypothesis testing for mean μ

	σ known	σ unknown	for σ
Point estimate	$\bar{X}$	$\bar{X}$	s
standard error(SE)	$\frac{\sigma}{\sqrt{n}}$	$\frac{s}{\sqrt{n}}$	
distribution	Normal	$t(n-1)$	$\chi^2(n-1)$
Test statistic	$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	$\frac{(n-1)s^2}{\sigma^2}$
Two tailed critical value	$\lvert z \rvert \geq z_{\alpha/2}$	$\lvert T \rvert \geq t_{\alpha/2}(n-1)$	$x > \chi^2_{\alpha/2}(n-1)$