PHP 2510 (Fall, 2009)

Third Midterm Exam

December 10

Name: ______________________________

Note: Write your name on the top of each page. There are three questions on this exam. Show your reasoning for each answer, and circle your final answer to avoid any ambiguities. Writing just the answer without explanation will not receive full credit, even if the answer is correct.

1. In a clinical trial, a new drug was compared to placebo for treatment of diabetic neuropathy where the positive response "+" of interest was improvement in peripheral sensory perception. Patients were assigned at random to receive one of the study treatments with 100 patients in each arm. Of these, 53 and 40 patients in each group, respectively, had a positive response.

response	new drug	placebo	
+	53	47	100
−	40	60	100

   (a) Estimate the risk difference (RD), risk ratio (RR) and odds ratio (OR). Interpret them.

   (b) Suppose that the sample size is sufficiently large that the central limit theorem (CLT) can be applied. Write out the distribution of RD implied by the CLT.

   (c) Is the statement "That RD = 0 implies that log(RR) = log(OR) = 0" true? Just answer Yes or No.

   (d) Using RD and for an $\alpha = 0.05$, test the hypotheses that H_0: the new drug is the same as placebo versus H_1: the new drug is effective, . (Given that $z_{0.05} = 1.64$, just state that your p-value is less than 0.05 or not). State your conclusion.

   (e) Compute the Pearson Chi-square statistic. Explain why it is for a two-sided test.

   (f) Graphically explain how to determine the p-value for the Pearson Chi-square statistic (No calculation is needed).

Name:

2. Suppose that a researcher wants to design a clinical trial to prove the efficacy of a new drug for treating a fatal disease. To alleviate ethical concerns, the researcher wants to implement an unequal arm design, where 80% of patients with the disease will be randomized to receive the new drug and the remaining 20% to placebo. Let n be the number of patients to receive placebo, and $4n$ be the number of patients to receive the new drug. Let X be the survival days of a treated patient, and Y be the survival days of an untreated patient (receiving placebo). Suppose both X and Y can be regarded as being normally distributed with means μ_X and μ_Y and a common variance σ^2.

   (a) Let $\bar{X}$ and $\bar{Y}$ be the sample means of the two arms. Write out the distributions of $\bar{X}$ and $\bar{Y}$. And, show that the distribution of $\bar{X} - \bar{Y}$ is

   $$\bar{X} - \bar{Y} \sim N(\mu_X - \mu_Y, 5\sigma^2/4n).$$

   (b) The hypotheses of interest are: H_0 $\mu_X - \mu_Y = 0$ (the new drug is not effective), versus H_1 $\mu_X - \mu_Y = d$ with $d > 0$. Show that

      i. For a significance level of α, the rejection region is $\bar{X} - \bar{Y} > c$ where

      $$c = z_\alpha \sqrt{5\sigma^2/4n}.$$

      ii. The power of the study is

      $$\Pr(Z > z_\alpha - d\sqrt{4n/5\sigma^2})$$

      where $Z \sim N(0, 1)$.

      iii. Therefore, to have a power of $(1 - \beta)$, the sample size for the placebo arm is

      $$n = \frac{5(z_\alpha - z_{1-\beta})^2\sigma^2}{4d^2},$$

      and the total number of patients for this trial is 5 times of this.

Name:

3. A study contains two arms, a treatment and a control arm. We are interested in the effect of the treatment as compared to the control on the outcome Y, which is normally distributed. Instead of calculating the two means and comparing their difference, we decide to fit a regression model. Suppose the study has n participants. We denote their arm assignments by X_i, $i = 1, 2, ..., n$, where if the ith individual is on the treatment, we let $X_i = 1$; otherwise if on the control, $X_i = 0$. The outcome of the ith individual is denoted by Y_i. A simple linear regression model is specified as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i.$$

Answer the following questions:

(a) State the three key statistical assumptions for simple linear models.

(b) i. Show that the model implies $E(Y|X = 0) = \beta_0$.

ii. Similarly, calculate what $E(Y|X = 1)$ and $\{E(Y|X = 1) - E(Y|X = 0)\}$ are.

iii. Based on (i) and (ii), explain what the parameter β_1 means in the context.

(c) The fitting results are

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x	4.092754	.915537	4.47	0.000	2.217362	5.968147
_cons	44.14149	.6254308	70.58	0.000	42.86036	45.42263

From this table, what are the estimates of β_0 and β_1? Summarize your findings from the study.

(d) Later, we find that *participant's age is not balanced between the two arms.* So we further include age (denoted by W) as an explanatory variable into the model, which now becomes

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + \epsilon_i.$$

The results are:

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
w	.5686275	.0423411	13.43	0.000	.4817508	.6555042
x	5.395219	.350131	15.41	0.000	4.676809	6.113628
_cons	28.14167	1.213342	23.19	0.000	25.65209	30.63124

i. Find the estimates of β_0, β_1 and β_2, and interpret them.

ii. Explain why the estimates of β_1 from the two models differ so much.

iii. Let $var(\epsilon) = \sigma^2$. An estimate of σ^2 is $\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i - \hat{\beta}_2 W_i)^2/(n - 3)$. Explain why the sum should be divided by $(n - 3)$.