Name: Section:

STAT 113, Spring 99 – Midterm 3

Note: On all problems, please show your work. Just the correct answer without justification and intermediate results is not acceptable.

1. In a survey of college students, it was found that $X = 69$ of the $n = 230$ students surveyed had at least one parent who had heart disease. Let p denote the population proportion of students with parents of whom at least one had had heart disease.

1a. [3pts] Give a 95% confidence interval $[a, b]$ for p, i.e., find a and b, such that

$$P(a \leq p \leq b) = 0.95.$$

Let $\hat{p} = X/n = 69/230 = 0.3$. Let $z_{\alpha/2} = 1.96$ denote the $(1 - \alpha/2)$ quantile of the standard normal distribution.

$$P(\underbrace{\hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}}_{=.3-1.96\sqrt{.3\cdot.7/230}=0.24} \leq p \leq \underbrace{\hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}}_{=.3+1.96\sqrt{.3\cdot.7/230}=0.36}) = 1 - \alpha \tag{1}$$

1b. [3pts] If we wanted the width $b - a$ of the confidence interval to be at most 0.02, how large a sample size n would we need?

$$(b - a) = 2\ z_{\alpha/2}\sqrt{\hat{p}\,(1-\hat{p})/n} = 0.02 \Rightarrow n \geq \left(2\,z_{\alpha/2}/0.02\right)^2 (\hat{p}\,(1-\hat{p})) = 8067.36.$$

Thus $n \geq 8068$.

1c. [2pts] Consider a hypthesis test

$$H_0:\ p = 0.2 \text{ vs. } H_1:\ p \neq 0.2,$$

with $\alpha = 0.05$ and using the test statistic $Y = (X - np)/\sqrt{\hat{p}\,(1-\hat{p})\,n}$.
What would be the conclusion of this test? Why?

The confidence interval (1) does not include $p = 0.2$. Therefore the test would reject H_0: $p = 0.2$.

2. Assume $x_1, \ldots, x_{14}$ is a random sample of size $n = 14$ from a $N(\mu, \sigma^2)$ distribution. Consider testing the hypothesis

$$H_0 : \sigma^2 = \sigma_0^2 \text{ vs. } H_1 : \sigma^2 > \sigma_0^2.$$

with $\sigma_0^2 = 1.0$. Use the test statistic

$$Y = \frac{(n-1)s_x^2}{\sigma_0^2}.$$

2a. Assume we observe $s_x^2 = 1.9$. Perform a hypothesis test for H_0 vs. H_1.

(i) [2pts] Give the distribution of Y under H_0
(state the name of the distribution and the relevant parameters).

$Y \sim \chi^2_{n-1}$, i.e., under H_0 the sampling distribution of Y is a χ^2 distribution with $(n-1)$ degrees of freedom.

(ii) [2pts] Fixing $\alpha = 0.05$, find the critical cutoff y^* for the test statistic Y.
I.e., find y^*, such that for $Y > y^*$ we reject H_0.

$y^* = 22.4$

(iii) [3pts] Find the p-value (observed significance level) for the observed $s_x^2 = 1.9$.

$$p = P(Y \geq \frac{(n-1)s_x^2}{\sigma^2}) = P(Y \geq \underbrace{\frac{13\ 1.9}{1.0}}_{24.7}) = 0.025$$

2b. [3pts] For $\sigma^2 = 3.2$, find the power of the test derived in problem **2a.**
Note: If you did not solve **2a.**, assume $y^* = 20$ (this is *not* the correct solution).

The power under $\sigma^2 = 3.2$ is the probability of *rejecting* H_0, computed under $\sigma^2 = 3.2$.

$$1 - \beta = P_{\sigma^2=4.5}(Y \geq 22.5) = P\left(\frac{(n-1)s^2}{\sigma_0^2} \cdot \frac{\sigma_0^2}{\sigma^2} \geq 22.4 \cdot \frac{\sigma_0^2}{\sigma^2}\right)$$
$$= P(X > 7.0) = 0.900,$$

Here X is a χ^2 distributed r.v. with $\nu = 13$ d.f.

2c*. [2pts] To achieve a power of $1 - \beta = 0.99$ for $\sigma^2 = 3.2$, how many observations (n) are required?

To achieve a power of 0.99 we need in the power calculation of **2d.**

$$22.4 \cdot \frac{\sigma_0^2}{\sigma^2} = 7.0$$

to be equal to the 0.99 quantile. From the χ^2 table we find $d.f. = 18$, i.e., $n = 19$.

3. An educational testing organization prepared a new version of an aptitude test. The average score on the old version, based on the scores of thousands of individuals who took the test, is 500. The new test is tried out on a sample of $n = 64$ students selected at random from those eligible to take the test; their average score is $\bar{x} = 480$, and their sample standard deviation is $s_x = 80$ (i.e., $s_x^2 = 6400$).

Does the evidence suggest that the new version is *harder* than the old?

3a. [2pts] State H_0 and H_1.

Let $\mu = E(x_i)$ denote the expected value of a score x_i on this aptitude test. The hypotheses are:

$$H_0\colon\ \mu = 500 \text{ vs. } H_1\colon\ \mu < 500.$$

3b. [2pts] Find an appropriate test statistic.

$$t = \frac{\bar{x} - 500}{s_x/\sqrt{n}}$$

3c. [2pts] Under H_0, what is the distribution of this test statistic?

t is t-distributed with $(n-1) = 63$ d.f. **or** t is (approximately) $N(0,1)$.

3d. [2pts] Using $\alpha = 5\%$, find appropriate cutoff(s) for the test statistic.

The $\alpha = 0.05$ quantile of the t-distribution with $\nu = 63$ d.f. is -1.67

or (when using approximate normality): $t^* = -1.64$.

3e. [2pts] What is your final decision (reject H_0; or fail to reject H_0?).

Find $t = (\bar{x} - 500)/(s/\sqrt{n}) = -20/(80/8) = -2.0$.
Since $t = -2.0$ is beyond the cutoff $t^* = -1.67$ we reject H_0.

4. Assume $x_i \sim N(0, 1/\theta)$, $i = 1, \ldots, n$, is a random sample of size n. I.e. $E(x_i) = 0$ and $Var(x_i) = 1/\theta$ and the x_i are independent.

Hint: Remember the normal p.d.f. for $x \sim N(\mu, \sigma^2)$ is

$$p(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

4a. [3pts] Obtain an estimator of θ using the method of moments;

Match the sample variance s_x^2 and the population variance $1/\theta$ to get

$$\hat{\theta} = 1/s_x^2.$$

4b. [2pts] Find the joint p.d.f. $p(x_1, x_2, \ldots, x_n)$.

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i) == (2\pi)^{-n/2}\, \theta^{n/2}\, \exp\left(-\theta/2 \sum_{i=1}^{n} x_i^2\right).$$

4c. [3pts] Obtain an estimator of θ the method of maximum likelihood.

Note that

$$\log p(x_1, \ldots, x_n) = -n/2\, \log(2\pi) + (n/2)\, \log(\theta) - \theta/2 \sum x_i^2.$$

and

$$\frac{d}{d\theta} \log p(x_1, \ldots, x_n) = (n/2)\, 1/\theta - (1/2) \sum x_i^2$$

$$\Rightarrow \hat{\theta} = 1/\left(\frac{1}{n} \sum x_i^2\right).$$