STA 2002 Final Exam

1 Multiple Choice Questions (26 points)

1. (4 points) Let $X_1, X_2, \ldots, X_n \sim N(\mu, 16^2)$ independently. Consider the critical region $|\frac{\bar{X}-100}{16/\sqrt{n}}| \geq z_{\alpha/2}$ for testing $H_0 : \mu = 100$ against $H_1 : \mu \neq 100$. (Here $z_{\alpha/2}$ is the value such that $P(Z \geq z_{\alpha/2}) = \alpha/2$ for $Z \sim N(0,1)$.) Determine for each of the following statements whether they are true or false.

   (a) **True / False** The critical region is a uniformly most powerful critical region of size α.

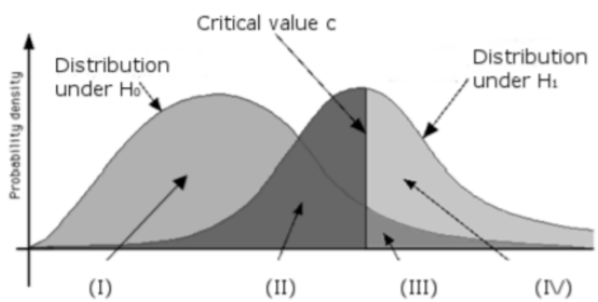
   (b) **True / False** A test with this critical region is a likelihood ratio test.

   (c) **True / False** There exists μ in H_1 for which the given critical region is not a best critical region.

   (d) **True / False** There exists μ in H_1 for which the given critical region is a best critical region.

   Solution: (a) False, (b) True, (c) True, (d) True.

2. (4 points) Suppose that a test for a simple null hypothesis H_0 against a simple alternative hypothesis H_1 will reject H_0 if the test statistic is greater than the critical value c. In the picture, the pdf of the test statistic under H_0 and H_1 is sketched, and four shaded areas are indicated by (I), (II), (III) and (IV). Choose which area represents each of the following:



   (a) Significance level : **I / II / III / IV**

   (b) Power: **I / II / III / IV**

   (c) P(type I error): **I / II / III / IV**

   (d) P(type II error): **I / II / III / IV**

   Solution: (a)III, (b) IV, (c) III, (d) II.

3. (6 points) Consider the 1-factor ANOVA model: Suppose $X_{ij} \sim N(\mu_i, \sigma^2)$ independently for $i = 1, 2, \ldots, a$ and $j = 1, 2, \ldots, b$, where $\mu_i = \mu + \alpha_i$, $\sum_{i=1}^{a} \alpha_i = 0$. Let $H_0 : \alpha_i = 0$ for all $i = 1, 2, \ldots, a$. Select for each of the following if it is an unbiased estimator for σ^2:

   (a) $\frac{SS(TO)}{ab-1}$ if H_0 is true? **yes / no**

   (b) $\frac{SS(TO)}{ab-1}$ if H_0 is false? **yes / no**

(c) $\frac{SS(E)}{ab-b}$ if H_0 is true? **yes / no**

(d) $\frac{SS(E)}{ab-b}$ if H_0 is false? **yes / no**

(e) $\frac{SS(A)}{a-1}$ if H_0 is true? **yes / no**

(f) $\frac{SS(A)}{a-1}$ if H_0 is true? **yes / no**

We recall that $SS(TO) = \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \bar{X}_{\cdot\cdot})^2$; $SS(E) = \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \bar{X}_{i\cdot})^2$; $SS(A) = \sum_{i=1}^{a}(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$; where $\bar{X}_{i\cdot} = \frac{1}{b}\sum_{j=1}^{b} X_{ij}$ for $i = 1, 2, \ldots, a$, and $\bar{X}_{\cdot\cdot} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} X_{ij}$.

Solution: (a) Yes, (b) No, (c) Yes, (d) Yes, (e) Yes, (f) No.

4. (4 points) Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n = 15$ from $N(0, \sigma^2)$, and let $C = \{(x_1, x_2, \ldots, x_{15})) : \sum_{i=1}^{15} x_i^2 \geq 100\}$ be a critical region for testing $H_0 : \sigma^2 = 4$ against $H_1 : \sigma^2 = 16$. Find the (approximate) power of the test.

   (a) 0.05

   (b) 0.95

   (c) 0.025

   (d) 0.975

   Hint: What is the distribution of $\sum_{i=1}^{15}\left(\frac{X_i - 0}{\sigma}\right)^2 = \sum_{i=1}^{15} X_i^2/\sigma^2$?

   Solution: (b) From the hint, $\sum_{i=1}^{15} X_i^2/\sigma^2 \sim \chi^2(15)$.

   The power is $P(\sum_{i=1}^{15} X_i^2 \geq 100; \sigma^2 = 16)$. Since under H_1, $\sum_{i=1}^{15} X_i^2/16 \sim \chi^2(15)$, the power is equal to $P(Q \geq 100/16 = 6.25)$, with $Q \sim \chi^2(15)$. Using Table IV, we find that $\chi^2_{0.975}(15) = 6.262$, i.e. $P(Q \geq 6.262) = 0.975$. So $P(Q \geq 6.25) \approx 0.975$.

5. (4 points) Let X and Y have a bivariate normal distribution with correlation coefficient ρ. To test $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$, a random sample of n pairs of observations is selected. Suppose that the sample correlation coefficient is $r = 0.72$. Using a significance level of $\alpha = 0.05$, find the smallest value of the sample size n so that H_0 is rejected.

   (a) 5

   (b) 6

   (c) 7

   (d) 8

   Solution: (b) $R \sim r(n-2), r_{0.025}(5) = 0.7544 > r > r_{0.025}(6) = 0.7067$

6. (4 points) Suppose that under H_0, a test statistic X follows a distribution with density function

$$f(x) = \begin{cases} \frac{1}{20}, & \text{if } 0 \leq x \leq 10 \\ \frac{1}{10}, & \text{if } 10 < x \leq 15 \\ 0, & \text{otherwise} \end{cases}$$

   Note that the mean of this distribution is $E(X) = 8.75$. Now, suppose we want to test $H_0 : E(X) = 8.75$ against alternative hypothesis $H_1 : E(X) < 8.75$. Given that the observed value of X is 2.5, what is the p-value?

(a) 0.8

(b) 0.25

(c) 0.1

(d) 0.125

Solution: (d) The p-value is $P(X < 2.5) = \int_0^{2.5} \frac{1}{20} dx = 0.125$.

2 Open Questions (74 points)

7. (10 points) Let X equal the weight in pounds of a "1-pound" bag of carrots. Let m equal the median weight of a population of these bags. We would like to test the null hypothesis $H_0 : m = 1.14$ against the alternative hypothesis $H_1 : m > 1.14$.

   (a) Assume the distribution of the weights of these bags is *not* normal, but that it does have a symmetric pdf. What type of test should we use? (Give the name of the test.)

   (b) If the observed weights (which have been ordered for your convenience) were

1.06	1.11	1.12	1.13	1.17
1.19	1.20	1.23	1.23	1.29

   compute the test statistic, and compare it to the critical value for $\alpha = 0.05$. What is your conclusion?

   Solution:

   (a) Wilcoxon signed rank test.

   (b) Computing the absolute deviations from $m_0 = 1.14$, we get

x_k	1.06	1.11	1.12	1.13	1.17	1.19	1.20	1.23	1.23	1.29
$\lvert x_k - m_0 \rvert$	0.08	0.03	0.02	0.01	0.03	0.05	0.06	0.09	0.09	0.15
signed rank	-7	-3.5	-2	-1	+3.5	5	6	8.5	8.5	10

   The test statistic is the sum of the signed ranks: $w = 28$. Under H_0, $W \sim N(0, \frac{10(11)(21)}{6})$, so we can compare $\frac{28-0}{\sqrt{385}} = 1.425$ to the critical value $z_{0.05} = 1.645$. Since $1.425 < 1.645$, we do not reject H_0.

8. (12 points) We are interested in estimating $\pi_{0.7}$, the 70-th percentile of weights of "80-pound" bags of water softener pellets, based on a random sample of $n = 14$ weights. Consider the interval $(x^{(8)}, x^{(13)})$, where $x^{(i)}$ is the i-th smallest of the $n = 14$ observations.

   (a) What is the confidence level of this interval for $\pi_{0.7}$? Give an exact expression as a binomial probability.

   (b) Use a normal approximation with correction for continuity to estimate the probability in (a).

   (c) We have the following random sample of $n = 14$ weights (which have been ordered for your convenience):

80.16	80.27	80.27	80.28	80.28	80.32	80.32
80.35	80.38	80.40	80.51	80.53	80.56	80.59

   Based on this data, give a point estimate for $\pi_{0.7}$, and the confidence interval $(x^{(8)}, x^{(13)})$.

   Solution:

(a) The confidence level is

$$P(X^{(8)} < \pi_{0.7} < X^{(13)}) = P(8 \le W \le 12)$$

where $W = \#\{i : X_i < \pi_{0.7}\}, W \sim$binom$(14, 0.7)$, i.e. it is $\sum_{k=8}^{12} \binom{14}{k} 0.7^k (1-0.7)^{14-k}$.

(b) We can approximate this probability by approximating $W \overset{approx}{\sim} N((0.7)(14), 0.7(1-0.7)14)$:

$$P(W \le 12) - P(W \le 7) \approx P(Z \le \frac{12 + \frac{1}{2} - (0.7)(14)}{\sqrt{0.7(1-0.7)14}}) - P(Z \le \frac{7.5 - (0.7)(14)}{\sqrt{0.7(1-0.7)14}})$$
$$= P(Z \le 1.575) - P(Z \le -1.341) = 0.94 - 0.09 = 0.85$$

(c) We estimate $\pi_{0.7}$ by $x^{0.7(14+1)} = x^{(10.5)} = \frac{1}{2}(x^{(10)} + x^{(11)}) = \frac{1}{2}(80.40 + 80.51) = 80.455$. The confidence interval is $(80.35, 80.59)$.

9. (10 points) Two men were randomly selected who had a smoking history classified as heavy, two men who had a moderate smoking history, and two men who had never smoked. The men in each category were randomly assigned to one of two stress tests: bicycle ergometer and a treadmill. The time until maximum oxygen uptake was recorded in minutes as follows:

	Stress Test	
Smoking History	Bicycle	Treadmill
Nonsmoker	12.5	17.4
Moderate	10.6	15.2
Heavy	8.5	12

To analyze this data, we computed an ANOVA table:

Source	Sum of Squares (SS)	Deg. of Freedom (DF)	Mean Square (MS)	F value
Smoking History (A)	22.21	___	___	___
Stress Test (B)	___	___	___	___
Error	0.543	___	___	
Total	50.92	___		

(You may want to start by filling in the underlined blanks).

(a) Perform a hypothesis test with significance level $\alpha = 0.01$ to test if smoking history affects the mean time until maximum oxygen uptake. Make sure to clearly state the test statistic, critical value or p-value for the test, and the conclusion.

(b) Perform a hypothesis test with significance level $\alpha = 0.01$ to test if the type of stress test affects the mean time until maximum oxygen uptake. Make sure to clearly state the critical value or p-value for the test, and the conclusion.

Solution: The full ANOVA Table is:

Source	Sum of Squares (SS)	Degrees of Freedom (DF)	Mean Square (MS)	F value
Smoking History	22.21	2	11.105	40.88
Stress Test	28.167	1	28.167	103.68
Error	0.543	2	0.272	
Total	50.92	5		

(a) Compare 40.88 to critical value $F_{2,2}(0.01) = 99$. The effect of smoking history is NOT significant at $\alpha = 0.01$.

(b) Compare 103.68 to critical value $F_{1,2}(0.01) = 98.5$. The effect of stress test IS significant at $\alpha = 0.01$.

10. (14 points) It has been claimed that, for a coin minted before 2000, the probability of observing heads when tossing the coin is 0.4. Let X denote the number of heads that occur when 2 coins are tossed at random. In 50 experiments we observed $x = 0, 1,$and 2 respectively $15, 20, 15$ times.

(a) Give a test statistic and test $H_0 : X \sim \text{bin}(2, 0.4)$ at $\alpha = 0.05$.

(b) Now, we want to test the null hypothesis that the distribution of X is $\text{bin}(2, p)$ for some $0 < p < 1$. Give a the method-of-moments estimate $\hat{p}$ for p based on the observations.

(c) Give a test statistic and test $H_0 : X \sim \text{bin}(2, p)$ at $\alpha = 0.05$

Solution:

	$x = 0$	$x = 1$	$x = 2$
Probability	0.36	0.48	0.16
Expected	18	24	8
Observed	15	20	15

(a)
$$q = \frac{(15-18)^2}{18} + \frac{(20-24)^2}{24} + \frac{(15-8)^2}{8} = 7.29 > 5.991 = \chi^2_{0.05}(3-1)$$

Reject the hypothesis that X is $\text{bin}(2, 0.4)$ at $\alpha = 0.05$ significance level.

(b) Since $\bar{x} = 1$, and $E[X] = 2p$, the method-of-moments estimator is $\hat{p} = 0.5$.

(c) With $\hat{p} = 0.5$,

	$x = 0$	$x = 1$	$x = 2$
Probability	0.25	0.5	0.25
Expected	12.5	25	12.5
Observed	15	20	15

$$q = \frac{(15-12.5)^2}{12.5} + \frac{(20-25)^2}{25} + \frac{(15-12.5)^2}{12.5} = 2 < 3.841 = \chi^2_{0.05}(3-1-1)$$

Do not reject the hypothesis that X is $\text{bin}(2, p)$ at $\alpha = 0.05$ significance level.

11. (12 points) Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from the normal distribution $N(\mu, \sigma_0^2)$, where σ_0^2 is known but μ is unknown.

(a) Find the likelihood ratio test for $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. Show that this critical region for a test with significance level α is given by $|\overline{X} - \mu_0| > z_{\alpha/2}\sigma_0/\sqrt{n}$.

(b) Test $H_0 : \mu = 59$ against $H_1 : \mu \neq 59$ when $\sigma^2 = 225$ and a sample of size $n = 100$ yielded $\bar{x} = 56.13$. Let $\alpha = 0.05$.

(c) What is the approximate p-value of this test? Note that H_1 is a two-sided alternative.

Solution:

(a)

$$\begin{aligned}
\lambda &= \frac{[1/(2\pi\sigma_0^2)]^{n/2}\exp[-\sum_{i=1}^n (x_i-\mu_0)^2/(2\sigma_0^2)]}{[1/(2\pi\sigma_0^2)]^{n/2}\exp[-\sum_{i=1}^n (x_i-\bar{x})^2/(2\sigma_0^2)]} \\
&= \exp\left[\frac{-\sum_{i=1}^n (x_i-\bar{x}+\bar{x}-\mu_0)^2}{2\sigma_0^2} + \frac{\sum_{i=1}^n (x_i-\bar{x})^2}{2\sigma_0^2}\right] \\
&= \exp\left[\frac{-n(\bar{x}-\mu_0)^2}{2\sigma_0^2}\right] \leq k
\end{aligned}$$

$$\begin{aligned}
\frac{-n(\bar{x}-\mu_0)^2}{2\sigma_0^2} &\leq \ln k \\
\frac{|\bar{x}-\mu_0|}{\sigma_0/\sqrt{n}} &\geq c \\
|z| = \frac{|\bar{x}-59|}{15/\sqrt{n}} &\geq z_{\alpha/2}
\end{aligned}$$

(b)

$$|z| = \frac{|56.13-59|}{15/10} = |-1.913| < 1.96 = z_{0.05/2}$$

Therefore, We don't reject H_0.

(c) $0.05 <$ p-value$= P(|Z| \geq 1.913) < 0.1$

12. (6 points) Suppose X is a discrete random variable, it has the following pmf:

x	-1	0	1
$p(x)$	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

where θ is an unknown parameter ($0 < \theta < 1$). A random sample yielded the following data: $x_1 = -1, x_2 = 0, x_3 = -1$. Find the value of the maximum likelihood estimator of θ.

Solution: $\frac{5}{6}$

The maximum likelihood function is

$$\begin{aligned}
L(\theta) &= \theta^2 \cdot 2\theta(1-\theta)\cdot\theta^2 = 2\theta^5(1-\theta) \\
\ln L(\theta) &= 5\ln\theta + \ln(1-\theta) + \ln 2 \\
\frac{d\ln L(\theta)}{d\theta} &= \frac{5}{\theta} - \frac{1}{1-\theta} \overset{\text{set}}{=} 0 \\
\theta &= \frac{5}{6}
\end{aligned}$$

13. (5 points)

Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from the Poisson distribution, $P(\lambda)$. Let $\overline{X}$ be the sample mean and and S^2 sample variance of the sample. Suppose that $\hat{\lambda} = a\overline{X} + (2-3a)S^2$ is an unbaised estimator of λ, find the value of a.
Hint: The cheat sheet includes the pdf, mgf, mean and variance of the Poisson distribution.

Solution: $\frac{1}{2}$
Since $X \sim P(\lambda)$, we get $E(X) = \mathrm{Var}(X) = \lambda$

$$\begin{aligned} E(\bar{X}) &= E(X) = \lambda, E(S^2) = \mathrm{Var}(X) = \lambda \\ E(\hat{\lambda}) &= \lambda \Longrightarrow a\lambda + (2-3a)\lambda = \lambda \\ \Longrightarrow & a = \frac{1}{2} \end{aligned}$$

14. (5 points) Let the pdf of X be defined by

$$f(x) = \begin{cases} \frac{2\theta^2}{(\theta^2-1)x^3}, & x \in (1,\theta), \\ 0, & \text{others} \end{cases}$$

Find the method-of-moments estimator of θ.

Solution: $\frac{\bar{X}}{2-\bar{X}}$

$$E(X) = \int_1^\theta x \cdot \frac{2\theta^2}{(\theta^2-1)x^3} dx = \frac{2\theta}{\theta+1}$$

Let $E(X) = \bar{X}$,that is $\frac{2\theta}{\theta+1} = \bar{X}$. Therefore, $\hat{\theta} = \frac{\bar{X}}{2-\bar{X}}$