MATH 476 Statistics

Fred J. Hickernell **Spring 2011**
Test 2 **Thursday, April 14**

Instructions:

i. *This test consists of* FOUR *questions. The scores of your* BEST THREE *attempts will count.*

ii. *The time allowed for this test is 75 minutes.*

iii. *This test is closed book, but you may use 2 double-sided letter-size sheets of notes.*

iv. *Calculators, even of the programmable variety, are allowed. Computers using JMP or MATLAB, are also allowed. No internet access, web browsing, email, chat, etc. is allowed.*

v. *Show all your work to justify your answers. Answers without adequate justification will not receive credit.*

1. (35 marks)
Let $X_1, \ldots, X_n$ be i.i.d. data with a Poisson distribution, i.e., the probability mass function of X_i is

$$f(x) = \frac{\theta^x e^{-\theta}}{x!}, \quad x = 0, 1, 2, \ldots$$

a) Let $Y = X_1 + \cdots + X_n$. Show that Y is a sufficient statistic for estimating θ.

Answer: The likelihood function is

$$\begin{aligned} L(x_1, \ldots, x_n|\theta) &= f(x_1) \cdots f(x_n) = \frac{\theta^{x_1} e^{-\theta}}{x_1!} \cdots \frac{\theta^{x_n} e^{-\theta}}{x_n!} \\ &= \frac{\theta^{x_1+\cdots+x_n} e^{-n\theta}}{x_1! \cdots x_n!} = \underbrace{\theta^y e^{-n\theta}}_{K_1(y,\theta)} \times \underbrace{\frac{1}{x_1! \cdots x_n!}}_{K_2(x_1,\ldots,x_n)}, \end{aligned}$$

where $y = x_1 + \cdots + x_n$. The factoring of the likelihood function establishes that Y is a sufficient statistic.

b) Derive the maximum likelihood estimator for θ.

Answer: The log likelihood is

$$\begin{aligned} \ell(\theta) &= \ln(L(x_1, \ldots, x_n|\theta)) = \ln(K_1(y, \theta)) + \ln(K_2(x_1, \ldots, x_n)) \\ &= y \ln(\theta) - n\theta + \ln(K_2(x_1, \ldots, x_n)) \end{aligned}$$

Taking the derivative with respect to θ and setting this derivative to zero implies

$$0 = \ell'(\theta) = \frac{y}{\theta} - n \implies \theta = y/n = \bar{x}$$

Since $\ell''(\theta) = -y/\theta^2 < 0$, it follows that $\bar{x} = (x_1 + \cdots + x_n)/n$ is the maximum likelihood estimator.

2. (35 marks)
Daniel, the Provost of CIT, claims that the Office of Technology Services (OTS) receives on average only 30 service tickets per day, and so the present OTS staff is sufficient to handle them. Matt, the Director of OTS, thinks that the Provost's number is too low. He wants to refute this claim, and will construct a hypothesis test to do so.

Matt notes that the number of service tickets each day are i.i.d. Poisson random variables with mean θ and variance θ. Thus, Y, the number of service tickets for 100 days, is a Poisson random variable with mean 100θ and variance 100θ. By the Central Limit Theorem Y is approximately a normal (Gaussian) random variable with mean 100θ and variance 100θ, i.e., $Y \approx N(100\theta, 100\theta)$.

a) What are the null hypothesis and alternative hypothesis for Matt's test?

Answer: $H_0 : \theta = 30$, $H_a : \theta > 30$.

b) Use the statistic $Y \approx N(100\theta, 100\theta)$ to construct a test for this hypothesis with Type I error of (approximately) 2.5%.

Answer: Since $Y \approx N(100\theta, 100\theta)$,

$$2.5\% \approx \text{Prob}\left(\frac{Y - 100\theta}{\sqrt{100\theta}} > 1.96\right) = \text{Prob}\left(Y > 100\theta + 1.96\sqrt{100\theta}\right).$$

For $\theta = 30$, $100\theta + 1.96\sqrt{100\theta} \approx 3107$. *Thus, if* $Y > 3107$, *the null hypothesis can be rejected.*

c) If Matt records 3150 service tickets in 100 days, can he refute Daniel's claim?

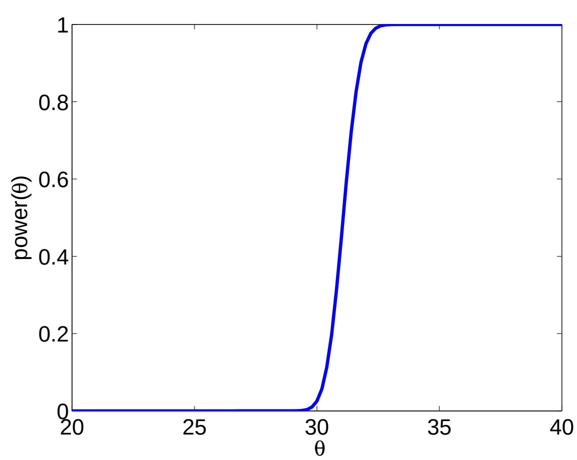
Answer: Yes. See above. However, even though Daniel's claim is refuted, he and Matt are still on good terms because Matt used proper statistical methods.

d) Sketch the power function for this test.

Answer: The power function is

$$\begin{aligned}\text{power}(\theta) &= \text{Prob}(\textit{Reject } H_0|\theta) = \text{Prob}(Y > 3107|\theta) \\ &= \text{Prob}\left(\frac{Y - 100\theta}{\sqrt{100\theta}} > \frac{3107 - 100\theta}{\sqrt{100\theta}}\Big|\theta\right) = 1 - \Phi\left(\frac{3107 - 100\theta}{\sqrt{100\theta}}\right)\end{aligned}$$

where Φ *is the cumulative distribution function of the standard normal random variable.*



3. (35 marks)
Jianing is evaluating three different manufacturing processes. Their yields in eight trials each are

A	28	32	29	31	32	26	35	30
B	26	24	33	29	29	28	30	27
C	30	28	28	32	34	33	31	36

Does there exist strong evidence at the 5% significance level that the three processes are different? What are Jianing's null and alternative hypotheses, what is the p-value, and what is her conclusion?

Answer: No there does not exist strong evidence at the 5% *significance level that the three processes are different. The null and alternative hypotheses are*

$$H_0 : \mu_A = \mu_B = \mu_C, \qquad H_a : \mu_A \neq \mu_B \text{ or } \mu_B \neq \mu_C \text{ or } \mu_C \neq \mu_A.$$

Using JMP to perform an ANOVA analysis, $p = 0.0812$*, which is not significant. Thus, one cannot tell if there is a difference.*

4. (35 marks)
Kayla is investigating whether the her company's training program is effective in improving the sales yield of its salespeople. The volume of sales in thousands of dollars generated by ten salespeople in the three months prior to the training and the three months after completing the training is recorded below:

Before	43	56	20	35	78	16	55	47	66	41
After	52	55	40	47	90	21	61	42	65	49

Does there exist strong evidence at the 5% significance level that the training produces an improvement in sales volume? What are Kayla's null and alternative hypotheses, what is the p-value, and what is her conclusion?

Answer: Yes there does exist strong evidence at the 5% *significance level that the training helps. The null and alternative hypotheses are*

$$H_0 : \mu_D = 0, \qquad H_a : \mu_D > 0,$$

where μ_D *is the mean of the difference in sales volume after and before the training. Using JMP to perform a Matched Pair analysis, the one-sided* p*-value is* 0.0110*, which is definitely significant.*