MATH 476 Statistics

Fred J. Hickernell | Test 2 | Thursday, April 9, 2026

Instructions:

i. *This test has* ***THREE*** *question(s). Attempt all. The maximum number of points is* ***100****.*
ii. *The time allowed is 75 minutes.*
iii. *This test is closed book, but you may use 4 double-sided letter-size sheets of notes.*
iv. *No calculators or other electronic devices are allowed. Phones must be placed in your bags under your desks or at the front of the room. Hands must be on top of your desks.*
v. *You are expected to simplify any expressions that can be simplified as whole numbers, e.g.,* $36/15 = 12/5$*, but you may leave other expressions as is, e.g.,* $\sqrt{47}$ *or* $18/29$*.*
vi. *You will be provided all of the critical values that you need.*
vii. *Show all your work to justify your answers. Answers without adequate justification will not receive credit.*
viii. *Print your name clearly on each answer sheet.*
ix. *Off-site students may contact the instructor as directed by your syllabus.*

I understand these instructions and have not relied on any help for this exam beyond what is allowed, nor provided any help to anyone else beyond what is allowed.

______________________________ ____________________
Signature *Date*

____
Printed Name

Upper quantiles: q_α **satisfies** $\mathbb{P}(X > q_\alpha) = \alpha$

X	α	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
$\text{Norm}(0,1)$	z_α	-2.58	-2.33	-1.96	-1.64	-1.28	1.28	1.64	1.96	2.33	2.58

---

1. (36 points) Determine whether or not each of the following statements is true. If a statement is false, correct it to make it true.

   a. (6 points) For an linear regression model with n data of the form $\boldsymbol{Y} = \mathsf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon} \sim \text{Norm}(\mathbf{0}, \sigma^2 \mathsf{I})$, the *true coefficient vector*, $\boldsymbol{\beta}$, is deterministic, but unknown.

   Answer: True. Since $\boldsymbol{\beta}$ *is unknown, we use regression to estimate it.*

   b. (6 points) For the linear regression model in part a., the width of the *confidence interval* for the mean response, $\mu(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\beta}$, tends to zero as $n \to \infty$.

   Answer: True. The confidence interval is proportional to $1/\sqrt{n}$*.*

c. (6 points) For the linear regression model in part a., the width of the *prediction interval* for a new observation, Y, tends to zero as the sample size, n, goes to ∞.

Answer: False. The width of the prediction interval is $2t_{n-d,\alpha/2}S\sqrt{1+\boldsymbol{x}^\top(\mathsf{X}^\top\mathsf{X})^{-1}\boldsymbol{x}}$, *which tends to* $2z_{\alpha/2}\sigma$.

d. (6 points) To test whether the mean of a distribution equals 10, use the hypotheses $H_0 : \overline{x} = 10$ and $H_A : \overline{x} \neq 10$.

Answer: False. Hypotheses must be written in terms of population parameters, *e.g.,* $H_0 : \mu = 10$ *and* $H_A : \mu \neq 10$.

e. (6 points) The Type I error of an hypothesis test is the probability of rejecting H_0 when it is actually true.

Answer: True. That is the definition.

f. (6 points) The power of an hypothesis test is computed based on the data collected.

Answer: False. The power of an hypothesis test only depends on the hypotheses, the distribution of the test statistic, and the Type I error. It does not depend on data.

2. (52 points) The city claims that its helpline answers phone calls with an average wait time of no more than 10 minutes or 600 seconds, but you are skeptical. Through a Freedom of Information Act (FOIA) request you find that in the recent past, 1600 calls have been received with an average wait time of $\overline{x} = 650$ seconds and a standard deviation of $s = 400$ seconds.

   a. (6 points) What should your null and alternative hypothesis be?

   Answer: Let $H_0 : \mu = 600$ *seconds and* $H_A : \mu > 600$ *seconds. One may also use* $H_0 : \mu \leq 600$.

   b. (25 points) Does this large data set provide compelling evidence to reject the null hypothesis at the $\alpha = 5\%$ significance level?

   Answer: Under H_0, *we have approximately* $Z = \frac{\overline{X}-\mu}{S/\sqrt{n}} \dot{\sim} \mathrm{Norm}(0,1)$, *but we observe*

$$z = \frac{\overline{x}-\mu}{s/\sqrt{n}} = \frac{650-600}{400/\sqrt{1600}} = \frac{50}{10} = 5,$$

   which is far beyond the $z_{0.05} = 1.64$ *critical value. We reject* H_0.
   Alternatively,

$$\begin{aligned} Z = \frac{\overline{X}-\mu}{S/\sqrt{n}} \dot{\sim} \mathrm{Norm}(0,1) \implies 5\% &\approx \mathbb{P}(Z \geq 1.64) = \mathbb{P}\left(\frac{\overline{X}-\mu}{S/\sqrt{n}} \geq 1.64\right) \\ &= \mathbb{P}\left(\overline{X} \geq \mu + 1.64S/\sqrt{n}\right) \\ &= \mathbb{P}\left(\overline{X} \geq 600 + 1.64 \times 400/\sqrt{1600}\right) \\ &= \mathbb{P}\left(\overline{X} \geq 600 + 16.4\right) = \mathbb{P}\left(\overline{X} \geq 616.4\right) \end{aligned}$$

Since $\overline{x} = 650$ is inside this rejection region, $\mathcal{R} = [616.4, \infty)$, there is reason to reject *the null hypothesis that $\mu = 600$.*

c. (6 points) What is the (approximate) rejection region, $\mathcal{R}$, for the *sample mean* for your hypothesis test?

Answer: See above: $\mathcal{R} = [616.4, \infty)$.

d. (15 points) How small can the sample size, n, be and still reject the null hypothesis if $\overline{x}$ and s are as given above?

Answer: Since

$$\mathcal{R} = \left[\mu + \frac{1.64S}{\sqrt{n}}, \infty\right) = \left[600 + \frac{1.64 \times 400}{\sqrt{n}}, \infty\right)$$

It follows that

$$\overline{x} = 650 \in \mathcal{R} \iff 650 \geq 600 + \frac{1.64 \times 400}{\sqrt{n}} \iff 50 \geq \frac{1.64 \times 400}{\sqrt{n}}$$
$$\iff \sqrt{n} \geq \frac{1.64 \times 400}{50} \iff n \geq (1.64 \times 8)^2 \approx 172$$

3. (12 points) Suppose that $X_1, \ldots, X_n \overset{\text{IID}}{\sim} \text{Bern}(p)$. Derive the rejection region, $\mathcal{R}$, for the likelihood ratio test for

$$H_0 : p = 0.5, \qquad H_A : p = 0.6.$$

You *do not* need to determine the cut-off value for a particular α, but just derive the form of $\mathcal{R}$. Show that the rejection region only depends on $T = X_1 + \cdots X_n$.

Answer: The likelihood and likelihood ratio in terms of the data, $\boldsymbol{X} = (X_1, \ldots, X_n)$, where $T = X_1 + \cdots + X_n$, are

$$L(p \mid \boldsymbol{X}) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i} = p^T(1-p)^{n-T}$$
$$\Lambda(\boldsymbol{X}) = \frac{L(0.5 \mid \boldsymbol{X})}{L(0.6 \mid \boldsymbol{X})} = \frac{0.5^T(0.5)^{n-T}}{0.6^T(0.4)^{n-T}} = \left(\frac{0.4}{0.6}\right)^T \times \left(\frac{0.5}{0.4}\right)^n = \left(\frac{2}{3}\right)^T \times \left(\frac{5}{4}\right)^n.$$

So the rejection region is

$$\mathcal{R} = \left\{T : \Lambda(\boldsymbol{X}) = \left(\frac{2}{3}\right)^T \times \left(\frac{5}{4}\right)^n \leq k\right\},$$

or equivalently

$$\mathcal{R} = \{T \geq c\},$$

for a suitably chosen c depending on α and n, since $(2/3)^T$ decreases as T increases, while $(5/4)^n$ is constant with respect to the data.

Test 2 Scores

Number of Students: 23, Minimum: 18, Maximum: 100, Mean: 69, Median: 74
Standard Deviation: 23.3, Quartiles (Q1, Q3): (54, 85)

Stem	Leaf
10	0
9	578
8	002258
7	0348
6	123
5	4
4	27
3	5
2	4
1	8