STAT 3332 Statistics for Life Sciences

Fall 2007 — Final Examination — 11/29/07

• Do **not** turn in the question sheets. **Only the answer sheet** will be accepted.
• There are **32** questions. Total points **100**: **3** per question, and **4** for *printing* name.
• Allotted time: **120** minutes.
• For each question, choose or give the **best answer**. Explanations (clearly indicating for which question) provided **neatly** on the front or back of the answer sheet may count for **partial credit** if needed.

---

1. The sample space of outcomes for 3 tosses of a coin is

$S = \{\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\}$.

For the two events $A = \{\text{HHH, HHT, THT}\}$ and $B = \{\text{THT,TTT}\}$ in S,

[(A)] $A \cap B = \{\text{THT}\}$ (B) $A \cup B = \emptyset$ (C) $\overline{A} = B$ (D) $\overline{B} = S$

---

2. (Continuation) Let $P(s)$ be any probability function on the outcomes s in S, and define the random variable $X(s)$ = number of Heads in outcome s. The induced probability $P(X = 2)$ is given by

[(A)] $P(\text{HHT}) + P(\text{HTH}) + P(\text{THH})$
(B) $P(\text{HTT}) + P(\text{HTH}) + P(\text{HHH})$
(C) $P(\text{HTT}) + P(\text{THT}) + P(\text{TTT})$

---

3. Let random variable X have a continuous probability distribution with density function $f(x)$. Then $P(-20 \le X \le -10) =$

(A) $f(-20) - f(-10)$ [(B)] $\int_{-20}^{-10} f(x)dx$ (C) $\int_{-\infty}^{\infty} f(x)dx$ (D) $f(-20)/f(-10)$

---

4. (Continuation) And $P(X = 0) =$

(A) $f(0)$ (B) 1 [(C)] 0 (D) 1/2

---

5. For the following data set of 9 values listed in increasing order, give the *median*.

−12 4 23 50 [600] 2,999 12,021 56,001 100,000

---

6. (Continuation) *Adding* the constant $c = 5$ to each value in this data set adds 5 to the

(A) median (B) mean (C) 3rd quartile [(D)] all of these

---

7. Suppose that an animal's genome has 30,000 genes, each independently having a very small chance $\frac{1}{10{,}000}$ of experiencing a mutation today. Then the mean number of mutations occurring today is

(A) $n + p = 30{,}003$ (B) $np^2 = 0.0003$ [(C)] $np = 3$

---

8. A *boxplot* uses the

(A) standard deviation (B) 10th percentile (C) 90th percentile [(D)] none of these

---

9. Let events A, B, and C have respective probabilities 1/2, 1/3, and 1/4, and let $A \cap B$, $B \cap C$, and $A \cap C$ have respective probabilities 1/5, 1/7, and 1/8. Which two events are *independent*?

(A) A and B (B) B and C [(C)] A and C

---

10. If 20 patients in a study report that they experienced angina in the past week, in how many ways can 3 different patients be chosen from this group for further examination this morning?

(A) $\frac{20}{3}$ [(B)] $\frac{20!}{3!17!}$ (C) $\frac{20!}{3!}$ (D) 20×3

---

11. Suppose that independent random samples from two distributions F and G are

From F: 8 20 21 27 42

From G: 19 26 28 60 61

For testing $H_0 : F = G$, the *Wilcoxon rank sum test statistic* W is

(A) $8 + 20 + 21 + 27 + 42$

[(B)] $1 + 3 + 4 + 6 + 8$

(C) $(1 + 3 + 4 + 6 + 8) - (2 + 5 + 7 + 9 + 10)$

---

12. In an opthalmology study of an eye drop, each participant randomly received either the active drug (A) or a placebo (P) in the left (L) eye and the other of (A) and (P) in the right (R) eye. Effectiveness was scored for each eye of each participant. To test the hypothesis of no difference between treatment and placebo, one may apply a one-sample test in this problem by computing for each participant the difference

(A) $L - R$ (B) $L - A$ (C) $L - P$ [(D)] $A - P$ (E) $A - R$

---

13. *Cardiovascular Disease.* Suppose for a population of persons it is known that

P(person is hypertensive) = .02,
P(person is normal) = .98,
P(chest pain | hypertensive) = .10,
P(chest pain | normal) = .05.

Then the probability that a randomly selected person both is normal and has chest pain is

(A) .10 + .05 = .15 [(B)] .98 × .05 = .049 (C) .98

---

14. (Continuation) The conditional probability P(hypertensive | chest pain) is

(A) $\frac{.98\times.05}{.98\times.05+.02\times.10}$ (B) $\frac{.98\times.02}{.98\times.02+.05\times.10}$ [(C)] $\frac{.10\times.02}{.10\times.02+.05\times.98}$

---

15. For the discrete probability distribution

x	0	1	2	3	4	5
$p(x)$	0.10	0.15	0.25	0.25	0.15	0.10

the *expected value* $E(X)$ is

(A) $\frac{1+2+3+4+5}{5}$

(B) $\frac{1\times0.15+2\times0.25+3\times0.25+4\times0.15+5\times0.10}{5}$

[(C)] $1 \times 0.15 + 2 \times 0.25 + 3 \times 0.25 + 4 \times 0.15 + 5 \times 0.10$

---

16. For sampling from any population with mean μ and variance σ^2, the standardized sample mean $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$

(A) has mean 0

(B) has variance 1

(C) has approximately a standard normal distribution

[(D)] satisfies each of (A), (B), (C)

(E) satisfies none of (A), (B), (C)

---

17. The probability distribution for the number of successes in 4 independent success-failure trials with success probability 1/4 is the binomial(4, 1/4) distribution:

x	0	1	2	3	4
$p(x)$	$\frac{81}{256}$	$\frac{108}{256}$	$\frac{54}{256}$	$\frac{12}{256}$	$\frac{1}{256}$

Given that *at least 2 successes occur* in the 4 trials, what is the conditional probability that the number of successes is *exactly 2*?

(A) 0 (B) $\frac{54}{256}$ [(C)] $\frac{54}{54+12+1}$ (D) $\frac{54+12+1}{256}$ (E) ∞

---

18. In hypothesis testing at level of significance $\alpha = .05$ using a p-value, the hypothesis H_0 would be rejected if

(A) p-value $= 0.235$ (B) p-value $= 0.072$ $\boxed{\text{(C)}}$ p-value $= 0.049$

---

19. Consider testing $H_0 : \mu = 133$ (my age at death, in years) versus $H_1 : \mu > 133$ with σ unknown, based on the sample mean $\overline{X}$ and the sample standard deviation s for a sample of size $n = 100$. To compute the p-value we need to use the H_0-distribution of $\frac{\overline{X}-133}{s/10}$. The needed distribution is

$\boxed{\text{(A)}}$ t with 99 degrees of freedom (B) $N(0, 1)$ (C) chi-square with 99 degrees of freedom

---

20. A hypothesis about $\mu_1 - \mu_2$ is to be tested based on $\overline{X}_1 - \overline{X}_2$ and sample variances s_1^2 and s_2^2 from independent samples of equal size n from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, with σ_1^2 and σ_2^2 unknown. A p-value would be based on the t distribution with degrees of freedom given by

(A) $2(n - 1)$ $\boxed{\text{(B)}}$ Satterthwaite's approximation (C) Normal approximation

---

21. For sampling of size 100 from $N(\mu, 25)$, with sample mean $\overline{X}$ and sample standard deviation s, a 95% confidence interval for μ is given by

(A) $\overline{X} \pm s$ (B) $\overline{X} \pm 1.96\,\frac{s}{10}$ (C) $\overline{X} \pm t_{99,0.975}\,\frac{s^2}{99}$ $\boxed{\text{(D)}}$ $\overline{X} \pm 1.96\,\frac{5}{10}$

---

22. For sampling of size 30 from any population with mean μ and variance σ^2, the sample variance $s^2 = (29)^{-1}\sum_{i=1}^{30}(X_i - \overline{X})^2$ has mean $E(s^2) =$

(A) μ (B) $\mu + 1.96\sigma$ (C) $\frac{\sigma^2}{\sqrt{30}}$ $\boxed{\text{(D)}}$ σ^2

---

23. Consider two independent samples of equal size n from populations with possibly differing means μ_1 and μ_2 but equal variance $\sigma^2 = 25$. Using the respective sample variances s_1^2 and s_2^2, the best estimator of σ^2 is $\hat{\sigma}^2 =$

$\boxed{\text{(A)}}$ $\frac{s_1^2+s_2^2}{2}$ (B) $\max\{s_1^2, s_2^2\}$ (C) $\frac{1}{s_1} + \frac{1}{s_2}$ (D) $\min\{s_1^2, s_2^2\}$ (E) $\sqrt{s_1^2 \times s_2^2}$

---

24. Consider *categorical data* in the 3×2 contingency table (with row totals, column totals, and the grand total shown in **bold**)

19	14	9	**42**
5	10	6	**21**
24	**24**	**15**	**63**

For the (1, 2) cell with *observed count* $O_{12} = 14$, the *estimated expected count* is $\widehat{E}_{12} =$

14 $\frac{24+42}{63}$ $\frac{24-42}{63}$ $\boxed{\frac{24\times42}{63}}$ 63

---

25. (Continuation) For testing homogeneity or independence, the test statistic $\sum_{cells}(O_{ij} - \widehat{E}_{ij})^2/\widehat{E}_{ij}$ has approximate distribution *chi square* with degrees of freedom (encircle)

$6 - 1 = 5$ $\boxed{(3 - 1) \times (2 - 1) = 2}$ 0 6 63

---

26. (Continuation) If the row and column totals are *random* with the total sample size (**63**) fixed, the statistic $\sum_{cells} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$ may be used for testing

(A) homogeneity of 3 populations $\boxed{\text{(B)}}$ independence of 2 variables (C) normality

---

27. Random numbers are used in statistics for

(A) selection of sample units from a population

(B) avoiding of bias in assigning patients to study groups

$\boxed{\text{(C)}}$ both of (A), (B)

(D) none of (A), (B)

28. The linear regression model for the relation between X and Y is described by

(A) $Y = \alpha + \beta X + \varepsilon$ with $E(\varepsilon) = 0$.

(B) $E(Y \mid X = x) = \alpha + \beta x$

[(C)] both of these

(D) neither of these

---

29. From a set of hematologic data for patients with aplastic anemia, a regression line relating number of lymphocytes (per mm^2) (Y) to % reticulocytes (X) was fitted:

$$Y = 2000 + 100X.$$

For two subjects whose % reticulocytes differ by 5, by how much do their numbers of lymphocytes differ, on the average?

(A) 0 [(B)] 500 (C) 1000 (D) 2500

---

30. The correlation between X and Y, $\rho_{XY} = \text{Cov}(X, Y)/\sigma_X\sigma_Y = E[(X - \mu_X)(Y - \mu_Y)]/\sigma_X\sigma_Y$,

(A) lies between -1 and $+1$

(B) measures linearity in the relationship

(C) measures dependence in the relationship

[(D)] each of (A), (B), and (C) is true

(E) none of (A), (B), and (C) is true

---

31. In fitting a linear regression model $E(Y \mid X = x) = \alpha + \beta x$ by least squares with data (X_i, Y_i), the estimates of α and β are chosen to minimize

(A) the sum of squared fitted values $\{\widehat{Y}_i, 1 \leq i \leq n\}$

[(B)] the sum of squared residuals $\{Y_i - \widehat{Y}_i, 1 \leq i \leq n\}$

(C) the sum of squared deviations $\{Y_i - \overline{Y}, 1 \leq i \leq n\}$

---

32. For measuring the goodness of fit of a linear regression model, one may use

(A) The square of the sample correlation

(B) The ratio of the sum of squares for regression to the total sum of squares

[(C)] Both (A) and (B)

(D) Neither (A) nor (B)

- When you are finished, *please hand in* **only** *the answer sheet.*
- *KEEP THE QUESTION SHEETS.*
- Please **depart quietly**, and *leave the vicinity before discussing the test.*

Thank you for your attention to these details.