STAT 3332 Statistics for Life Sciences

Spring 2007 — Final Examination — 4/30/07

• Do **not** turn in the question sheets. **Only the answer sheet** will be accepted.
• There are **32** questions, with **3** points per question plus **4** points for neatness of name and signature on the answer sheet. Total points: **100**.
• Allotted time: **120** minutes.
• For each question, choose the **best answer**. Explanations for any question (clearly indicate which) may be provided on the margin or back of the answer sheet and may count for **partial credit** in the case of a wrong answer.

---

1. The sample space of outcomes for 2 tosses of a coin is $S = \{\text{HH, HT, TH, TT}\}$. For the two events $A = \{\text{HH, HT}\}$ and $B = \{\text{TH,TT}\}$ in S,

(A) $A \cap B = S$ (B) $A \cup B = \emptyset$ * (C) $\overline{A} = B$ (D) $\overline{B} = S$

---

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

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

---

3. (Continuation) And $P(X = -10) =$

(A) $f(-10)$ (B) $-10/20$ * (C) 0 (D) 1/2

---

4. For the sample space $S = \{\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\}$ and probability function $P(s)$ on outcomes s in S, define the random variable $X(s) =$ number of Heads in s. The induced probability $P(X = 1)$ is given by

(A) $P(\text{HHH}) + P(\text{HTT}) + P(\text{TTT})$
(B) $P(\text{HHT}) + P(\text{HTT}) + P(\text{THH})$
* (C) $P(\text{HTT}) + P(\text{THT}) + P(\text{TTH})$

---

5. For the following data set of 10 values listed in increasing order, give the *median*. (Answer: 550)

-120 -12 4 23 500 600 999 12,021 56,001 100,000

---

6. (Continuation) *Adding* the constant $c = 15$ to each value in this data set adds 15 to the

(A) variance (B) interquartile range (C) standard deviation * (D) none 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 number of mutations occurring today is *binomial* with $n = 30{,}000$ and $p = \frac{1}{10{,}000}$. Approximation of this distribution by *Poisson* with mean 3 is justified because

(A) $n + p = 30,003$ * (B) np^2 is small (C) both of these

---

8. A *boxplot* uses the

(A) standard deviation (B) 10th percentile * (C) interquartile range (D) each 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 1/6, 1/12, and 1/16. Which two events are *dependent*?

(A) A and B (B) B and C * (C) A and C

---

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

(A) $10/3 = 3.333$ * (B) $\binom{10}{3} = 240$ (C) $10! = 3{,}628{,}800$ (D) $10 \times 3 = 30$

---

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

From F:	8	20	21	27	42			
From G:	0	19	26	28	60	61	63	66

For testing $H_0 : F = G$, the *Wilcoxon rank sum test statistic* W is (encircle best answer)

(A) $8 + 20 + 21 + 27 + 42 = 118$

* (B) $2 + 4 + 5 + 7 + 9 = 27$

(C) $1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 + 11 + 12 + 13 + 14 = \frac{14\times15}{2} = 105$

---

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 these in the right (R) eye. Effectiveness was scored for each eye of each participant. To best reduce the problem to a one-sample test, compute for each participant

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

---

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 person) = .10,
P(chest pain | normal person) = .05.

Then the probability that a randomly selected person is normal is

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

---

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

* (A) $\frac{.049}{.98\times.05+.02\times.10} = 0.96$ (B) .002 (C) .049

---

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) 1 * (B) 2.5 (C) 4

---

16. For sampling from any population with mean μ and variance σ^2,

(A) the sample mean $\overline{X}$ has mean 0

(B) the sample mean $\overline{X}$ has variance 1

* (C) the standardized sample mean $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ has mean 0 and variance 1

(D) each of (A), (B), (C) is true

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

---

17. The probability distribution for the number of successes in 3 independent success-failure trials with success probability 1/3 is the binomial$(3, 1/3)$ distribution:

x	0	1	2	3
$p(x)$	$\frac{8}{27}$	$\frac{12}{27}$	$\frac{6}{27}$	$\frac{1}{27}$

Given that *at least 2 successes occur* in the 3 trials, what is the conditional probability that the number of successes is *exactly 2*?

(A) 0 (B) $\frac{2}{7}$ (C) $\frac{6}{27}$ * (D) $\frac{6}{7}$ (E) ∞

---

18. In hypothesis testing at level of significance $\alpha = .01$ using a p-value, the hypothesis H_0 would be rejected if the p-value for the data satisfied

* (A) p-value = 0.005 (B) p-value > 0.01 (C) p-value > 0.05

---

19. Consider testing $H_0 : \mu = 33$ (my age) versus $H_1 : \mu > 33$ with $\sigma = 10$ known, based on the sample mean $\overline{X}$ for a sample of size $n = 130$. To compute the p-value we need to use the distribution of $\frac{\overline{X}-33}{10/\sqrt{130}}$. This is

(A) t with 99 degrees of freedom * (B) $N(0, 1)$ (C) chi-square

---

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 $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, with σ^2 unknown. A p-value would be based on the distribution

* (A) t with $2(n-1)$ degrees of freedom (B) $N(0, 1)$ (C) chi-square

---

21. For sampling of size 30 from $N(\mu, \sigma^2)$ with both parameters unknown, a 90% confidence interval for μ based on the sample mean $\overline{X}$ and the sample standard deviation s is

(A) $\overline{X} \pm s$ (B) $\overline{X} \pm 1.96\, s$ (C) $\overline{X} \pm t_{29,0.30} \frac{s^2}{29}$ * (D) $\overline{X} \pm t_{29,0.95} \frac{s}{\sqrt{30}}$

---

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) σ^2 (C) $\frac{\sigma^2}{\sqrt{30}}$ (D) 31

---

23. Consider two independent samples of sizes n_1 and n_2, respectively, 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 $\hat{\sigma}^2 =$

(A) $\frac{s_1^2+s_2^2}{2}$ (B) $\max\{s_1^2, s_2^2\}$ * (C) $\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ (D) $\frac{1}{s_1} + \frac{1}{s_2}$

---

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

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

For the $(1, 1)$ cell with *observed count* $O_{11} = 19$, the *expected count* is $E_{11} =$

(A) 19 (B) $\frac{24+42}{63}$ (C) $\frac{24-42}{63}$ (D) 63 * (E) $\frac{24\times42}{63}$

---

25. (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 * (B) independence of 2 variables (C) normality

---

26. Random numbers are used in statistics for

(A) sample selection (B) avoiding bias * (C) both of these

---

27. Among 40 randomly selected households in a neighborhood, 15 ($\hat{p} = 0.375$) favor letting Wal-Mart put a store in the neighborhood. To form a confidence interval for the actual proportion p of households favoring this action, approximate the exact distribution of $\hat{p}$ by

(A) $N(0, 1)$ (B) $N(0.375, 1)$ (C) $N(0.375, 0.375)$ * (D) $N\left(0.375, \frac{0.375 \times 0.625}{40}\right)$

---

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 2%, by how much do their numbers of lymphocytes differ, on the average?

(A) 100 * (B) 200 (C) 2000 (D) 2200

---

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

(A) normality * (B) linearity (C) exponentiality (D) binomiality

between X and Y.

---

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

(A) $\sum_{i=1}^{n} (\alpha Y_i - \beta X_i)^2$ (B) $\sum_{i=1}^{n} (Y_i - X_i)^2$ * (C) $\sum_{i=1}^{n} [Y_i - (\alpha + \beta X_i)]^2$

---

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)

- 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.