

Statistics Proficiency Exam Study Guide

This guide provides a general overview of statistics. It is not meant to be inclusive of all items included on the actual proficiency exam. You should refer to the textbook from your previous statistics class for a more comprehensive review of the material your course covered.

INFORMATION ABOUT QUALIFYING EXAM FOR STAT-UB 1

Reviewed 18 August 2015

Passing this qualifying exam allows the student to skip the four-credit course STAT-UB 1. The student must still take the two-credit course STAT-UB 3 on regression analysis.

A student who does not pass this qualifying exam will take the six-credit combined course STAT-UB 103.

This exam requires no serious calculations, but students may bring four-function (**non-graphing**) calculators if they wish. The exam is in the “closed book” format.

There are 20 multiple-choice problems on this test. There is no penalty for wrong answers, so it is recommended that you answer every question.

The exam stresses concepts rather than pure calculation.

Here are some sample questions. These are similar in spirit and difficulty to those on the actual exam. Solutions follow on page 8.

S1. Suppose that you flip a fair coin six times. Then $P(\text{two or fewer heads})$ is

(a) $\frac{6}{64}$

(d) $\frac{21}{64}$

(b) $\frac{7}{64}$

(e) $\frac{22}{64}$

(c) $\frac{15}{64}$

S6. Which of the following statements is true?

- (a) The median of a set of numbers is always larger than the mean.
- (b) The upper quartile of a set of numbers is less than or equal to the median of the set of numbers.
- (c) Negative values of the standard deviation indicate that the set of values is even less dispersed than would be expected by chance alone.
- (d) The standard deviation is commonly used because of its ease of calculation using only paper and pencil.
- (e) The mean is used because it is a commonly-understood simple statistical summary of a set of numbers.

S7. Steve used a sample of 26 values from a population which was assumed to follow the normal distribution. He tested the null hypothesis $H_0: \mu = 300$ versus alternative $H_1: \mu \neq 300$. Here μ is assumed to be the population mean. Using a significance level $\alpha = 0.05$, he rejected H_0 . This means that

- (a) If he took another sample of 26 values he would again reject H_0 .
- (b) The probability that H_0 is true is less than or equal to 0.05.
- (c) Steve may have committed a Type II error.
- (d) Steve needs a larger sample size.
- (e) If H_0 is really true, then the probability of rejecting it is at most 0.05.

S8. Based on a sample of 50 values from a population which is assumed to follow the normal distribution, Alicia computed a 95% confidence interval for the population mean μ , and she obtained (136.0 hr, 156.0 hr). Here “hr” means “hours,” the units of measurement for this situation. Based on the *same* set of 50 values, Brenda wanted to test the null hypothesis $H_0: \mu = 150$ hr against alternative $H_1: \mu \neq 150$ hr. Please choose one of the following statements.

- (a) If Brenda uses $\alpha = 0.05$ as her level of significance, then she will accept H_0 .
- (b) If Brenda uses $\alpha = 0.01$ as her level of significance, then she will reject H_0 .
- (c) Brenda’s result cannot be determined from the information provided.
- (d) If Brenda accepts H_0 , then she may be committing a Type I error.
- (e) Since the problem asks whether μ is 150 hours, Brenda needs to ask for more time to complete the work.

S9. At a novelty shop on Sixth Avenue, Eddie purchased an eight-sided die. This object has eight equal-sized triangular faces, and these faces have the numbers 1 through 8. Suppose that Eddie rolls the die ten times. What is the probability that the face showing the number 3 comes up at least once?

- (a) 0.1^8
- (b) $\left(\frac{1}{8}\right)^{10}$
- (c) $1 - \left(\frac{1}{8}\right)^{10}$
- (d) $1 - \left(\frac{7}{8}\right)^{10}$
- (e) $10 \times \left(\frac{1}{8}\right)^1 \times \left(\frac{7}{8}\right)^9$

S10. The total retail value of furniture sold on any day at Outdoor Furnishings is approximately normally distributed with mean \$12,000 and with standard deviation \$1,000. The symbol \bar{X} represents the average retail value over a four day period. The probability that \bar{X} exceeds \$11,500 is about

- (a) 0.90
- (b) 0.84
- (c) 0.75
- (d) 0.69
- (e) 0.91

S11. Stephanie flips four ordinary coins, and she notes the number of heads that result from this process. Find the conditional probability

P(exactly three heads | at least three heads)

- (a) $\frac{4}{16} = \frac{1}{4}$
- (b) $\frac{5}{16}$
- (c) $\frac{1}{5}$
- (d) $\frac{4}{5}$
- (e) 0

S12. The symbol \tilde{X} denotes the median of a sample of seven values. Which of the following statements is correct?

- (a) Exactly three values are below \tilde{X} and exactly three values are above \tilde{X} .
- (b) There is at least one value below \tilde{X} and at least one value above \tilde{X} .
- (c) There are at least four values $\leq \tilde{X}$.
- (d) Exactly four values are $\geq \tilde{X}$.
- (e) Exactly four values are $\leq \tilde{X}$.

S13. At the Cool De-Lite ice cream shop, the banana split is made with three scoops of ice cream, and the customers are allowed to select different flavors. If the shop has ten different flavors, and if you'd like to choose a three-flavor banana split, how many choices do you have?

- (a) 720
- (b) 360
- (c) 480
- (d) 120
- (e) 27

S14. Only one of the following probability statements about events A , B , and C is *not* correct. Which one?

- (a) $P[A \cup B \cup C] \geq P[A \cap B]$
- (b) $P[A \cup B \cup C] \leq P[A \cap B \cap C]$
- (c) $P[A \cup B] \geq P[A \cap C]$
- (d) $P[A \cup (B \cap C)] \leq P[A \cup B]$
- (e) $P[A] + P[B] + P[C] \geq P[A \cup B \cup C]$

S15. Sam collected a sample x_1, x_2, \dots, x_{32} of 32 values and computed the conventional 95% confidence interval for the unknown population mean μ , getting the interval 117.4 ± 26.0 . Which of the following statements is correct?

- (a) A 90% confidence interval for μ would have been slightly shorter.
- (b) If Sam had known the value of the population standard deviation σ , he could have given a shorter interval.
- (c) In the hypothesis testing framework $H_0: \mu = 130$ versus $H_1: \mu \neq 130$, the null hypothesis would have been rejected at significance level 0.05.
- (d) With a sample size larger than 32, the probability that the 95% confidence interval covers the true value of μ would be larger.
- (e) The population mean is 117.4.

S16. You and Frank are playing a game involving a sequence of coin tosses. You win the game as soon as the sequence shows two consecutive heads, and Frank wins the game as soon as the sequence shows tails-then-heads. For example, the sequence HH is a winner for you, and the sequence THH is a winner for Frank. The probability that you will win this game is

- (a) $\frac{1}{2}$
- (b) $\frac{3}{4}$
- (c) $\frac{1}{4}$
- (d) $\frac{2}{3}$
- (e) The probability cannot be determined until the first flip is observed.

S17. If you are testing the null hypothesis H_0 versus alternative H_1 there is the possibility of error. Which of the following is correct?

- (a) You must make either a Type I error or a Type II error.
- (b) If you fix the probability of Type I error at 0.05, then the probability of Type II error must be less than 0.05.
- (c) If you fix the probability of Type I error at 0.05, then the probability of Type II error must be greater than or equal to 0.05.
- (d) The probability of Type I error plus the probability of Type II error adds to 100%.
- (e) None of the above.

S18. Tiffany has been playing a casino game in which the probability of winning is 0.36. If she plays 100 times, then

- (a) the expected number of times she wins is 50, and the standard deviation of the number of times she wins is 5.
- (b) the expected number of times she wins is 36, and the standard deviation of the number of times she wins is 48.
- (c) the expected number of times she wins is 36, and the standard deviation of the number of times she wins is 4.8.
- (d) the expected number of times she wins is 0.36, and the standard deviation of the number of times she wins is 0.48.
- (e) the expected number of times she wins is 0.36, and the standard deviation of the number of times she wins is 0.048.

S19. If two cards are selected from a standard deck of 52 (with the first card's *not* being replaced before the second is selected), the probability that there will be one heart and one diamond is

- (a) $\frac{13}{52} \times \frac{13}{52}$
- (b) $\frac{13}{52} \times \frac{13}{51}$
- (c) $\frac{13}{52} \times \frac{12}{51}$
- (d) $2 \times \frac{13}{52} \times \frac{13}{51}$
- (e) $\frac{1}{2} \times \frac{13}{52} \times \frac{13}{51}$

20. Consider events G and H with $P(G) = 0.40$, $P(H) = 0.25$, $P(G \cup H) = 0.55$. Then events G and H are

- (a) independent.
- (b) mutually exclusive.
- (c) neither independent nor mutually exclusive.
- (d) complementary.
- (e) undefined.

SOLUTIONS:

S1. Choice (e) is correct. This experiment involves binomial probabilities, and we can calculate the probability of two or fewer heads as $P(\text{no heads}) + P(\text{one head}) + P(\text{two heads})$. This comes down to

$$\binom{6}{0}\left(\frac{1}{2}\right)^6 + \binom{6}{1}\left(\frac{1}{2}\right)^6 + \binom{6}{2}\left(\frac{1}{2}\right)^6 = \frac{1}{64} + \frac{6}{64} + \frac{15}{64}$$

The sum is $\frac{22}{64}$, and this is choice (e).

S2. The expected value is computed as

$$\begin{aligned} & 0 \times 0.83 + 5 \times 0.08 + 10 \times 0.05 + 15 \times 0.03 + 20 \times 0.01 \\ &= 0 + 0.40 + 0.50 + 0.45 + 0.20 \end{aligned}$$

The sum is 1.55, which is choice (d).

S3. Choice (d) is best. Here the sample average \bar{X} has a standard deviation of $\frac{20 \text{ grams}}{\sqrt{100}} = 2$ grams. The comparison point, 122 grams, is exactly one standard deviation above average in the distribution of \bar{X} , meaning $\frac{122 - 120}{2} = 1$. The normal distribution has about 68% of its probability within one standard deviation of the average. This leaves 16% of the probability above (mean + one standard deviation) and 16% below (mean - one standard deviation). Here the closest answer is 15%, choice (d).

S4. If all values are doubled, then both \bar{x} and s will double; this is choice (a).

S5. The correct interpretation of a confidence interval is (e).

S6. The correct statement is (e).

Statement (a) cannot possibly be true; for any list with median $>$ mean, we can simply multiply each value by -1 to produce a list with median $<$ mean.

The upper quartile corresponds to the 75th percentile, and it cannot be less than the median, which corresponds to the 50th percentile. Thus (b) is false.

As for (c), standard deviations are *never* negative, except by calculation error!

Statement (d) is false, as the standard deviation is not very easy to compute.

S7. The correct solution is (e).

Statement (a) involves a second sample. There is simply no basis for asserting that the two samples will agree.

Statement (b) relates to a common misconception. It gives the conditional probability $P(H_0 \text{ true} \mid \text{Steve's data}) \leq 0.05$. There is nothing in this structure which deals with the probability that H_0 is true, either before Steve collects his data or afterward. The proper interpretation is $P(\text{Steve's data} \mid H_0 \text{ true}) \leq 0.05$, and this is response (e).

Type II error happens when a false H_0 is accepted. Since Steve rejected H_0 , statement (c) is not even in play.

If Steve rejected H_0 with a sample of 26, he has enough data (at least for hypothesis testing purposes). Thus (d) is off the mark.

S8. The correct solution is (a). There is a direct correspondence between the confidence interval and this hypothesis test. Specifically, H_0 will be accepted at level $\alpha = 0.05$ while using the t test based on $t = \sqrt{50} \frac{\bar{X} - 150 \text{ hr}}{s}$ if and only if the value 150 hr is inside the 95% confidence interval $\bar{X} \pm t_{0.025;49} \frac{s}{\sqrt{50}}$.

Since 150 hr is inside the interval, H_0 will be accepted at level $\alpha = 0.05$. It follows that H_0 would also be accepted at the more stringent $\alpha = 0.01$, so response (b) must be incorrect.

Response (c) is incorrect; we are indeed able to state Brenda's result.

Since Type I error is rejecting H_0 when it is true, this error cannot be done by someone who accepts H_0 . Thus (d) is incorrect.

Response (e) is pure nonsense. The times in hours in the data base have no relationship to the times required by Alicia and Brenda.

S9. The correct solution is (d). This is constructed as $P(\text{at least one "3"}) = 1 - P(\text{no "3"}) = 1 - [P(\text{not 3 on one roll})]^{10} = 1 - \left(\frac{7}{8}\right)^{10}$. Response (e) is the probability of *exactly* one "3."

S10. The correct response is (b). Since \bar{X} is the average of four values, it will have a standard deviation $\frac{\sigma}{\sqrt{n}} = \frac{\$1,000}{\sqrt{4}} = \$500$. The problem is solved directly by standardizing:

$$P[\bar{X} > \$11,500] = P\left[\frac{\bar{X} - \$12,000}{\$500} > \frac{\$11,500 - \$12,000}{\$500}\right] = P[Z > -1.0]$$

where Z is the generic standard normal random variable. An exact use of the normal table would provide $P[Z > -1.0] = 0.8413$. The problem was stated as "approximately normally distributed," so 0.84 is the clear choice.

A common error comes in skipping the \sqrt{n} step, and working with the incorrect value $SD(\bar{X}) = \$1,000$. This would lead to $P[Z > -0.5]$, producing a value very close to 0.69.

S11. The correct solution is (d). Using the probability law of the binomial, we find

$$P(3 \text{ heads}) = \binom{4}{3} \times \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{2}\right)^1 = \frac{4}{16}$$

$$P(4 \text{ heads}) = \binom{4}{4} \times \left(\frac{1}{2}\right)^4 \times \left(\frac{1}{2}\right)^0 = \frac{1}{16}$$

Then $P(\text{at least 3 heads}) = P(3 \text{ heads}) + P(4 \text{ heads}) = \frac{5}{16}$.

Finally, we obtain the conditional probability as

$$\begin{aligned} & P(\text{exactly three heads} \mid \text{at least three heads}) \\ &= \frac{P(\text{exactly three heads} \cap \text{at least three heads})}{P(\text{at least three heads})} \\ &= \frac{P(\text{exactly three heads})}{P(\text{at least three heads})} = \frac{\frac{4}{16}}{\frac{5}{16}} = \frac{4}{5} \end{aligned}$$

S12. The correct solution is (c). The cause for concern is that there may be ties among the data values. Consider this set of seven values: $\{16, 16, 16, 16, 16, 23, 85\}$. The median is 16. Each of statements (a), (b), (d), and (e) is incorrect for this set. These statements, by the way, are also true:

- (f) There are at least four values $\geq \tilde{X}$.
- (g) There is at least one value exactly equal to \tilde{X} .

S13. The correct solution is (d). This is a “combinations” problem, and the number of choices is $\binom{10}{3} = \frac{10!}{3! 7!} = \frac{10 \times 9 \times 8}{3 \times 2 \times 1} = 120$.

S14. Only statement (b) is not correct.

In statements (a) and (c), the left side is a superset of A , and the right side is a subset of A . Certainly the left side has a larger probability.

The same property is true of statement (b), so the relationship should also be \geq in that case.

In statement (d), $(B \cap C)$ is a smaller set than B , so the right side must have a larger probability.

Statement (e) is a common fact about probabilities, though it's more often seen in the simple form $P[A \cup B] \leq P[A] + P[B]$.

S15. Statement (a) is correct.

Statement (b) seems to be asking you to compare the conventional interval

$\bar{x} \pm t_{0.025; n-1} \frac{s}{\sqrt{n}}$ to the normal interval $\bar{x} \pm z_{0.025} \frac{\sigma}{\sqrt{n}}$. It is guaranteed that

$t_{0.025; n-1} < z_{0.025}$, but the difference is very small. The relative sizes of these intervals depend primarily on whether s (the sample standard deviation) is bigger or smaller than σ (the population standard deviation). The s -versus- σ comparison could go either way.

The null hypothesis $H_0: \mu = 130$ is accepted if and only if the comparison value 130 is inside the confidence interval. Here 130 is clearly in the interval, so statement (c) is incorrect.

Statement (d) is pure nonsense. The probability that a 95% confidence interval covers the unknown parameter is designed to be 95%, no matter what the sample size.

The value 117.4 is \bar{x} , the sample mean. It's perhaps close to the population mean μ , but the two quantities are certainly not equal.

S16. The correct solution is (c). You will win if the sequence starts HH . For any other start, Frank must win. Thus, you win with probability $P[HH] = \frac{1}{4}$.

Part (e) reflects a basic confusion about probabilities. Certainly it can be shown that $P[\text{you win} \mid \text{first flip is } H] = \frac{1}{2}$ and $P[\text{you win} \mid \text{first flip is } T] = 0$, however the description "you win" is a well-defined event and certainly $P[\text{you win}]$ can be given as a number.

S17. The correct response is (e).

If H_0 is true, then you will either make a correct decision (by accepting H_0) or you will make a Type I error (by rejecting H_0). If H_1 is true, then you will either make a correct decision (by rejecting H_0) or you will make a Type II error (by accepting H_0). Whether H_0 is true or not true, you have the possibility of a correct decision. Thus statement (a) is wrong is stating that you must make an error.

The probability of Type II error is related to statistical design issues, especially sample size, and the probability of this error could be anywhere between 0 and 1. Thus statements (b), (c), and (d) are incorrect.

S18. The correct response is (c). The number of times she wins is a binomial phenomenon with $n = 100$ and $p = 0.36$. The expected number of times that she wins is $np = 100 \times 0.36 = 36$. The standard deviation of the number of times that she wins is $\sqrt{np(1-p)} = \sqrt{100 \times 0.36 \times 0.64} = 10 \times 0.6 \times 0.8 = 4.8$.

S19. The correct solution is (d). This is the usual solution:

$$\begin{aligned} P[\heartsuit \text{ first, } \diamondsuit \text{ second}] &= P[\heartsuit \text{ first}] \times P[\diamondsuit \text{ second} \mid \heartsuit \text{ first}] \\ &= \frac{13}{52} \times \frac{13}{51} \end{aligned}$$

The event can also occur in the mutually exclusive style [\diamondsuit first, \heartsuit second], which has the same probability. Thus the solution is $2 \times \frac{13}{52} \times \frac{13}{51}$

S20. The solution is (a). Observe that

$$\begin{aligned} P(G \cup H) &= 0.55 = P(G) + P(H) - P(G \cap H) \\ &= 0.40 + 0.25 - P(G \cap H) = 0.65 - P(G \cap H) \end{aligned}$$

It follows that $P(G \cap H) = 0.10$. Since this is exactly $P(G) \times P(H)$, these events are independent.