AP Statistics Exam Tips
2012 Edition

Daren Starnes
The Lawrenceville School
dstarnes@lawrenceville.org

BFW Publishers

May 3, 2012

#1. Get Prepared!

- Do released AP Exam MC and FR questions
  - Look at FR scoring rubrics from AP Central
- Do practice questions from other sources
  - Stats 4 Stem website
  - AP Statistics Review books
  - Cumulative AP Practice Tests in TPS 4e
- Review Formulas and Tables for AP Exam
- Review inference procedures
  - Larry Green's website
  - Inference summary from TPS 4e endpapers

#2: Be a smart test taker!

- General
  - Get a good night's sleep
  - Review TPS 4e Appendix A *AP Exam Tips*
  - Eat a healthy lunch
  - Check that your calculator is working
- Multiple choice section
  - Answer all the questions (no guessing penalty)
  - Cover up answers while reading the stem
- Free response section
  - Don't run out of time before Question 6!!
  - Start with questions you feel confident about
  - Use bullets and outlines rather than complete sentences whenever possible

#3: Use technology wisely

- Don't type in data just because it's there!
- On inference questions, use STAT TESTS for calculations. Only show formulas with values subbed in for partial credit if you get the same answer!
- Calculator speak = no full credit

(1) Binompdf(12,.2,3) = no full credit

Better: $P(X = 3) = (12C3)(0.2)^3(0.8)^9$

Minimal: binompdf with n = 12, p = .2, k = 3

(2) Normalcdf(90,105,100,5) = no full credit

Better: Find z score. Then draw, label, shade Normal curve

Minimal: normalcdf with lower bound = 90, upper bound = 1-5, mean = 100, std. dev. = 5

#4: Follow the 4-step inference process

How to Organize a Statistical Problem: A Four-Step Process		
	Confidence intervals (CIs)	**Significance tests**
STATE:	What *parameter* do you want to estimate, and at what confidence level?	What *hypotheses* do you want to test, and at what significance level? Define any *parameters* you use.
PLAN:	Choose the appropriate inference *method*. Check *conditions*.	Choose the appropriate inference *method*. Check *conditions*.
DO:	If the conditions are met, perform *calculations*.	If the conditions are met, perform *calculations*. • Compute the **test statistic.** • Find the ***P*-value**.
CONCLUDE:	*Interpret* your interval in the context of the problem.	*Interpret* the result of your test in the context of the problem.

Starting in 2010, partial credit is available on 4-step rubrics.

#4: Follow the 4-step inference process

2011 Question #4: High Cholesterol

High cholesterol levels in people can be reduced by exercise, diet, and medication. Twenty middle-aged males with cholesterol readings between 220 and 240 milligrams per deciliter (mg/dL) of blood were randomly selected from the population of such male patients at a large local hospital. Ten of the 20 males were randomly assigned to group A, advised on appropriate exercise and diet, and also received a placebo. The other 10 males were assigned to group B, received the same advice on appropriate exercise and diet, but received a drug intended to reduce cholesterol instead of a placebo. After three months, posttreatment cholesterol readings were taken for all 20 males and compared to pretreatment cholesterol readings. The tables below give the reduction in cholesterol level (pretreatment reading minus posttreatment reading) for each male in the study.

Group A (placebo)

Reduction (in mg/dL)	2	19	8	4	12	8	17	7	24	1

Mean Reduction: 10.20 Standard Deviation of Reductions: 7.66

Group B (cholesterol drug)

Reduction (in mg/dL)	30	19	18	17	20	−4	23	10	9	22

Mean Reduction: 16.40 Standard Deviation of Reductions: 9.40

Do the data provide convincing evidence, at the $\alpha = 0.01$ level, that the cholesterol drug is effective in producing a reduction in mean cholesterol level beyond that produced by exercise and diet?

Remember: Don't just start typing in the data unless you have a reason to do so!!

#4: Follow the 4-step inference process

2011 Question #4: High Cholesterol

Step 1: States a correct pair of hypotheses.

Let μ_A represent the mean cholesterol reduction if all such male patients at this hospital are advised on appropriate exercise and diet and also receive a placebo.

Let μ_B represent the mean cholesterol reduction if all such male patients at this hospital are advised on appropriate exercise and diet but receive the drug instead of a placebo.

The hypotheses to be tested are $H_0: \mu_A = \mu_B$ versus $H_a: \mu_A < \mu_B$.

Step 2: Identifies a correct test procedure (by name or by formula) and checks appropriate conditions.

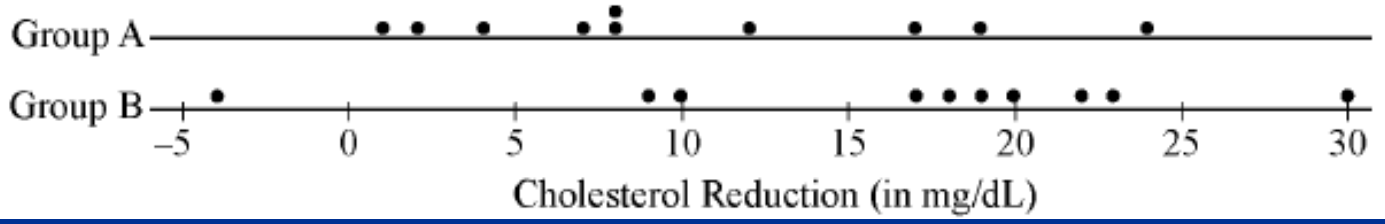
The appropriate procedure is a two-sample *t*-test.

When comparing two experimental treatments using a two-sample *t*-test, the subjects must be randomly assigned to the treatments. This condition is stated in the question (10 men were randomly assigned to group A and the remaining 10 men to group B).

#4: Follow the 4-step inference process

2011 Question #4: High Cholesterol

The second condition is that the two populations are approximately normally distributed or the sample sizes are sufficiently large. Because of the small sample sizes (10 in each treatment group), we need to check whether it is reasonable to assume that the samples came from populations that are normally distributed. The following dotplots reveal slight skewness and a possible outlier for group B, but it appears reasonable to proceed with the two-sample *t*-test.



Step 3: Demonstrates correct mechanics, including the value of the test statistic and *p*-value (or the rejection region).

The test statistic is: $t = \dfrac{\bar{x}_A - \bar{x}_B}{\sqrt{\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}}} = \dfrac{10.20 - 16.40}{\sqrt{\dfrac{7.66^2}{10} + \dfrac{9.40^2}{10}}} \approx -1.62$

With df = 17.3, *p*-value ≈ 0.062.

#4: Follow the 4-step inference process

2011 Question #4: High Cholesterol

Step 4: States a correct conclusion in the context of the problem, using the result of the statistical test.

Because the p-value is greater than the significance level of $\alpha = 0.01$, we fail to reject H_0. The data do not provide enough evidence at the 0.01 level of significance to conclude that the drug is effective in producing a mean cholesterol reduction beyond that provided by exercise and dietary advice.

- Don't "accept" the null hypothesis!!

#5: Don't shoot yourself in the foot!

- Misuse of term or notation → mandatory deduction
  - Putting statistics instead of parameters in hypotheses

    $H_0 : \bar{x}_{BR} = \bar{x}_{FF}$
  - Saying there's confounding when there isn't
  - Using experiment language in an observational study

    Blocking vs. stratified sampling; experiment vs. survey
- Advice:
  - Only use terms and symbols you know
  - It's better to explain in your own words than to use a term incorrectly
  - If you're unsure about notation, use words instead of symbols

#6: Naked answer = no credit

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

2009, Question #2

A tire manufacturer designed a new tread pattern for its all-weather tires. Repeated tests were conducted on cars of approximately the same weight traveling at 60 miles per hour. The tests showed that the new tread pattern enables the cars to stop completely in an average distance of 125 feet with a standard deviation of 6.5 feet and that the stopping distances are approximately normally distributed.

(a) What is the 70th percentile of the distribution of stopping distances?

(b) What is the probability that at least 2 cars out of 5 randomly selected cars in the study will stop in a distance that is greater than the distance calculated in part (a) ?

(c) What is the probability that a randomly selected sample of 5 cars in the study will have a mean stopping distance of at least 130 feet?

#6: Naked answer = no credit

2009, Question #2(a): 70th percentile of stopping distances

Part (a) is scored as follows:

Essentially correct (E) if the student clearly indicates which distribution is being used, along with the parameters (μ_X and σ_X), and correctly calculates the percentile value with appropriate justification (except for minor arithmetic or transcription errors). There are three components: distribution, parameters, and calculation of distance.

Notes:

- The standard notation $N(125,\ 6.5)$ defines distribution and parameters. Also, the *z*-score formula setup implies distribution and parameters. This applies only in part (a), because approximate normality is given in the stem of the problem.
- If the calculator command invNorm(0.70, 125, 6.5) is provided along with 128.4 feet *AND* an appropriately labeled sketch of a normal distribution is supplied, then the response should be scored as essentially correct (E). An appropriately labeled sketch must include correct labels for center and spread.
- If the calculator command invNorm(0.70, $\mu = 125$, $\sigma = 6.5$) is provided along with 128.4 feet, then the response should be scored as essentially correct (E).

Partially correct (P) if the student correctly supplies only two out of the three components.

Note: If the calculator command invNorm(0.70, 125, 6.5) is provided along with 128.4 feet, then the response should be scored as partially correct (P).

Incorrect (I) if the student correctly supplies at most one of the components.

#7: If you can't answer part of a question, don't assume you can't answer other parts

2010 Question #4

An automobile company wants to learn about customer satisfaction among the owners of five specific car models. Large sales volumes have been recorded for three of the models, but the other two models were recently introduced so their sales volumes are smaller. The number of new cars sold in the last six months for each of the models is shown in the table below.

Car Model	A	B	C	D	E	Total
Number of new cars sold in the last six months	112,338	96,174	83,241	3,278	2,323	297,354

The company can obtain a list of all individuals who purchased new cars in the last six months for each of the five models shown in the table. The company wants to sample 2,000 of these owners.

(a) For simple random samples of 2,000 new car owners, what is the expected number of owners of model E and the standard deviation of the number of owners of model E?

(b) When selecting a simple random sample of 2,000 new car owners, how likely is it that fewer than 12 owners of model E would be included in the sample? Justify your answer.

(c) The company is concerned that a simple random sample of 2,000 owners would include fewer than 12 owners of model D or fewer than 12 owners of model E. Briefly describe a sampling method for randomly selecting 2,000 owners that will ensure at least 12 owners will be selected for each of the 5 car models.

If needed, make up a reasonable answer for the part you can't answer…

#8: Know what distribution you're talking about!

2010 Question #2

A local radio station plays 40 rock-and-roll songs during each 4-hour show. The program director at the station needs to know the total amount of airtime for the 40 songs so that time can also be programmed during the show for news and advertisements. The distribution of the lengths of rock-and-roll songs, in minutes, is roughly symmetric with a mean length of 3.9 minutes and a standard deviation of 1.1 minutes.

(a) Describe the sampling distribution of the sample mean song lengths for random samples of 40 rock-and-roll songs.

(b) If the program manager schedules 80 minutes of news and advertisements for the 4-hour (240-minute) show, only 160 minutes are available for music. Approximately what is the probability that the total amount of time needed to play 40 randomly selected rock-and-roll songs exceeds the available airtime?

- Distribution of sample data (lengths of 40 songs)
- Population distribution (lengths of all rock-and-roll songs the station plays)
- Sampling distribution of statistic (e.g. sample mean song length)

#9: Don’t write too much

- Answer the question, then **shut up☺!**
- Space provided is more than enough
- The mandatory deduction rule
- The parallel solutions rule

#10:RTQ & ATQ!!

2011 Question #1: The Combine

1. A professional sports team evaluates potential players for a certain position based on two main characteristics, speed and strength.

   (a) Speed is measured by the time required to run a distance of 40 yards, with smaller times indicating more desirable (faster) speeds. From previous speed data for all players in this position, the times to run 40 yards have a mean of 4.60 seconds and a standard deviation of 0.15 seconds, with a minimum time of 4.40 seconds, as shown in the table below.

	Mean	Standard Deviation	Minimum
Time to run 40 yards	4.60 seconds	0.15 seconds	4.40 seconds

   Based on the relationship between the mean, standard deviation, and minimum time, is it reasonable to believe that the distribution of 40-yard running times is approximately normal? Explain.

#10:RTQ & ATQ!!

2011 Question #1: The Combine

(b) Strength is measured by the amount of weight lifted, with more weight indicating more desirable (greater) strength. From previous strength data for all players in this position, the amount of weight lifted has a mean of 310 pounds and a standard deviation of 25 pounds, as shown in the table below.

	Mean	Standard Deviation
Amount of weight lifted	310 pounds	25 pounds

Calculate and interpret the z-score for a player in this position who can lift a weight of 370 pounds.

AP Statistics Exam Tips

Recap

#1: Get prepared!

#2: Be a smart test taker.

#3: Use technology wisely

#4: Follow the 4-step inference process

#5: Don’t shoot yourself in the foot!

#6: Naked answer = no credit

#7: If you can’t answer part of a question, don’t assume you can’t answer other parts

#8: Know what distribution you’re talking about!

#9: Don’t write too much!

#10:RTQ & ATQ!!

Questions?

Email me:
dstarnes@lawrenceville.org