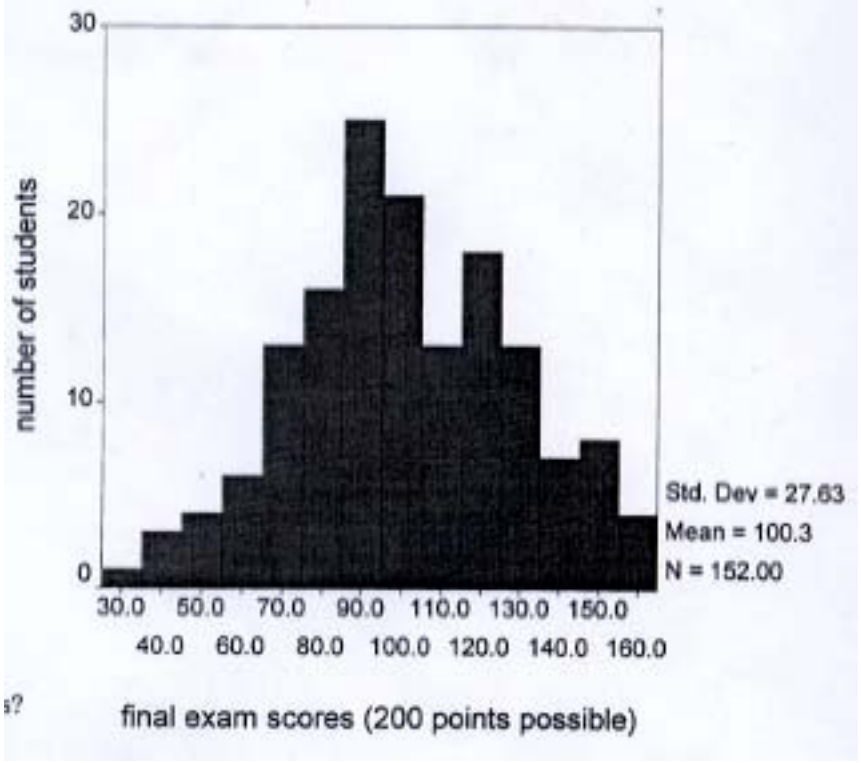
Introductory Statistics – A Self-Assessment Test (Answers are in boxes.)

Descriptive statistics on one variable

1. A class of 160 sociology students took a final exam, and the following graph shows the results of those who finished the test. What is this kind of graph called? A histogram or frequency distribution.



2. What do the heights of the bars represent in this graph?
Number of students scoring at each value of the variable.

3. What does N=152 mean?
Number of scores or people with scores.

4. What was the mode on this test? 90

5. What shape does the distribution of this variable have (approximately)?
Almost normal, slightly heavy on the high end.

6. The mean of all these final exam scores is 100.3. How was this number calculated?
All the scores added up, divided by 152.

7. What does the mean tell us about this set of scores?
Where the center of the distribution is.

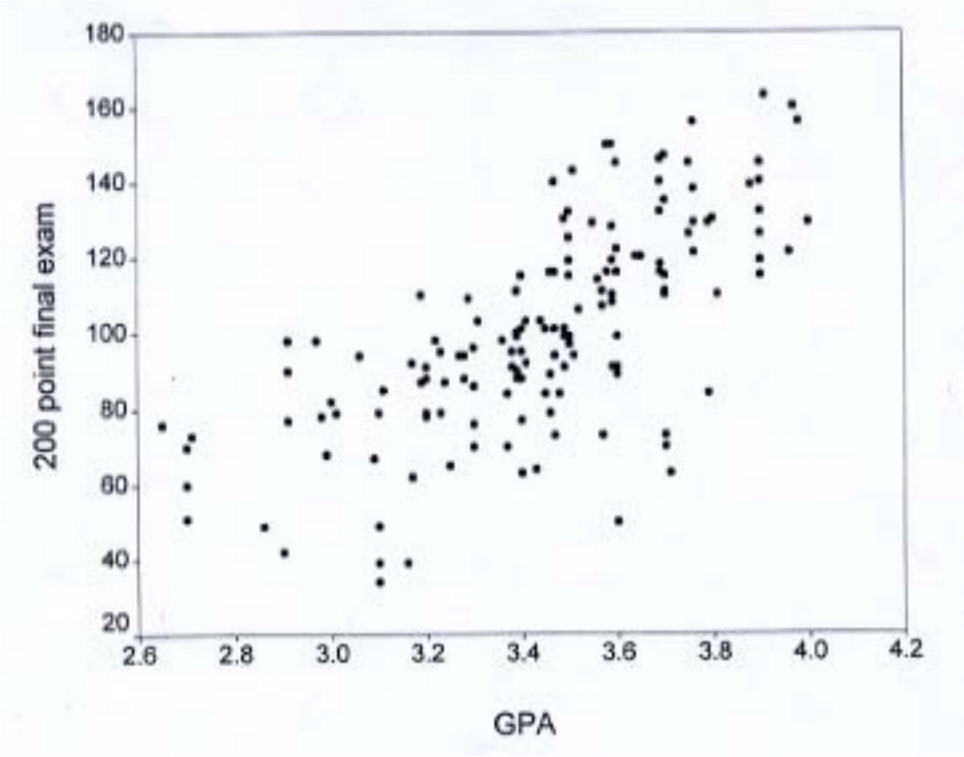
8. What does the abbreviation Std. Dev. stand for?
Standard Deviation, the variability of the scores.

9. In this example, what does the 27.63 tell us about these scores?
The typical deviation of all the scores around the mean is 27.63 points.

10. What is the approximate range of these scores? Range = 160-30 = 130 points.

Relationships between two variables

11. The following graph shows the relationship between the final exam scores of these sociology students and their overall GPAs. What is this kind of graph called? A scatterplot or scattergraph.

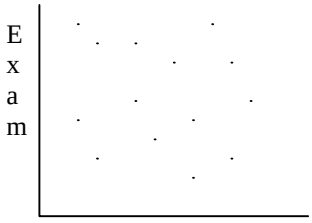


12. What does each dot on the graph represent?
Each individual's scores on both variables, the GPA and the final exam.

13. What type of relationship exists between final exam scores and GPA?
Positive correlation.

14. Suppose you made a graph showing the final exam scores and the students' heights in inches. We can safely assume these variables should be unrelated to each other. Describe or sketch approximately what this graph would look like.

A scatterplot would have a swarm of points going straight across, with no downward or upward trend.



15. What statistic quantifies (turns into a single number) the relationship between variables like these?
Pearson R.

Testing Yourself? You should have gotten at least 9 of 10 questions correct on page 1 and 4 of 5 correct on page 2.

Inferential Statistics

16. What is the difference between a population and a sample?

Population is all the individuals of interest to the researcher and to whom conclusions will be generalized. The sample is the subset of the population, drawn for researcher purposes, from which data are actually collected.

17. Suppose you wanted to see if men scored differently than women on the final exam in this sociology class. What kind of statistical analysis would you run? Two independent samples t-test.

18. The results of an appropriate analysis are shown below. What null hypothesis is being tested?

That the mean scores of men and women are the same, or that the sample of men and women came from the same population.

Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
200 point final exam	male	56	104.5000	25.6146	3.4229
	female	86	100.4651	27.4414	2.9591

	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						Lower	Upper
Equal variances assumed	.879	140	.381	4.0349	4.5913	-5.0424	13.1122
Equal variances not assumed	.892	123.349	.374	4.0349	4.5246	-4.9211	12.9909

19. In this analysis, what part does the "t" value play?

t is the test statistic comparing the means of the two groups. The larger it is, the less likely that the null hypothesis is true. When it reaches a critical value, the null is rejected.

20. What part does the "Sig. (2-tailed)" play?

This is the probability that the null hypothesis would be rejected when it is actually true (Type I error). If this probability is less than .05, we reject the null.

21. What conclusions could you draw about the difference between men's and women's scores?

There is a slight tendency for men (104.5) to score higher than women (100.5), but this 4-point difference is not significant. It probably is a random difference only.

22. If you wanted to estimate the mean of all the college women who have taken this test, what would be the best point estimate to use?

The women's sample mean, 100.4651.

23. If you were told that the 95% Confidence Interval for the women's mean was $94.48 < \mu < 106.45$, what would this indicate?

You can be 95% confident that the mean of the population of women's scores is within this interval.

Testing yourself? You should have gotten at least 4 out of 8 questions correct on this page.

24. Suppose you wanted to know if there was a relationship between sex (male or female) and whether students had a high school sociology course (yes or no). You make up a table such as the following. What is the size of the sample? 143

25. What does the number 48 mean in the center of the table? Number of women who had H.S. sociology.

sex * high school sociology Crosstabulation

Count

		high school sociology		
		no	yes	Total
sex	male	34	22	56
	female	39	48	87
Total		73	70	143

26. If the chi-square statistic for these data turned out to be 3.44 and the significance, .064, what conclusion would you draw?

Although there seems to be a tendency for a larger proportion of women to have had HS sociology, this trend is not quite significant.

Testing yourself? You should have gotten at least 2 of 3 questions correct on this page.