Typical Q's

1. A researcher reports that, on average, the participants in his study lost 10.4 pounds after two months on his new diet. A friend of yours comments that she tried the diet for two months and lost no weight, so clearly the report must be a fraud. Which of the following statements is correct?

This is the correct explanation of the situation

a. The report gives only the average. This does not imply that all participants in the study lost 10.4 pounds or even that all participants lost weight. Your friend's experience does not necessarily contradict the study results.

b. In order for the study to be correct, we must now add your friend's results to those of the study and compute a new average.

c. Your friend must not have followed the diet correctly because she did not lose weight.

d. Because your friend did not lose weight, the report must not be correct.

e. Not enough information to answer the question

---

2. You record the age, marital status and earned income of a sample of 1463 women. The number of variables you have recorded is

a. 1463 ← number of cases only.

b. four--age, marital status and income and number of women

quantitative

c. three--(age, marital status and income)

qualitative

d. 3 times 1463

e. two--age and income; marital status is not a variable because it doesn't have a unit like dollars or years

3. A consumer group surveyed the prices for a certain item in five different stores and reported the median price as $15. We visited four of the five stores and found the prices to be $10, $15, $15, and $25. Assuming that the consumer group is correct, the price of the item at the store that we did not visit

a. must be $15

b. must be above $15

c. must be below $15

d. can be any value

e. Not enough information to answer the question

9, 10, 15, 15, 25

10, 11, 15, 15, 25

10, 15, 15, 15, 25

10, 15, 15, 18, 25

10, 15, 15, 25, 30

No matter how we insert a number in this sequence, it doesn't change the fact the median is 15.

---

4. The average salary of all female workers is $35,000. the average salary of all male workers is $41,000. What must be true about the average salary of all workers?

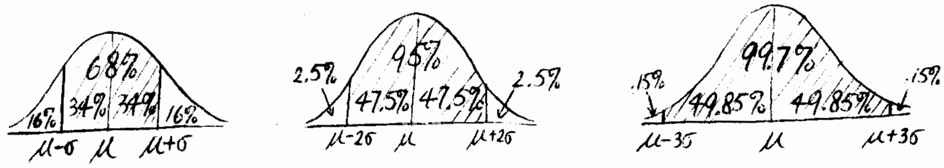
a It must be $38,000 ← only true when there are an equal number of male and female workers.

b. It must be larger than the median salary

c. It could be any number between $35,000 and $41,000 ← always true

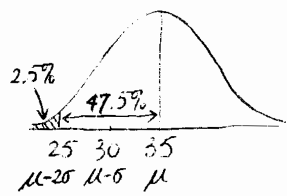
d. It must be larger than $38,000

e. Not enough information to answer the question



5 - 7. The administration of a college thinks that students will stay in school if they have good teachers. The teacher evaluation instrument rates teachers on a 50-point scale. The mean score is 35 with a standard deviation of 5. The scores are normally distributed. Use what you know about the 68%/95%/99.7% rule to answer this question.

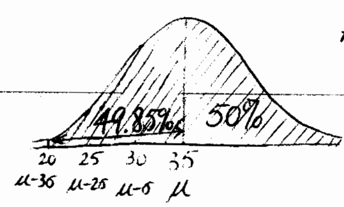
a. Find the probability that a teacher will score less than 25 on the teacher evaluation instrument.



Answer: 2.5%

b. Find the probability that a teacher will score more than a 20 on the teacher evaluation instrument.

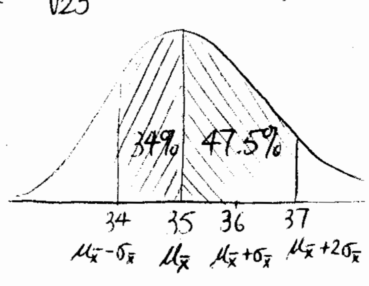
Answer: 50% + 49.85% = 99.85%



c. A sample of 25 teachers is selected. Find the probability that the mean of these 25 teachers is between 34 and 37 on the teacher evaluation instrument.

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{25}} = 1 \qquad \mu_{\bar{x}} = \mu$

This problem is about the sampling distribution of 25 teachers. While the $\mu_{\bar{x}}$ is same as the μ, the $\sigma_{\bar{x}} = 1$, not 5.



Answer: 34% + 47.5% = 81.5%

8. The heights of American men aged 18 to 24 are approximately normally distributed with mean 68 inches and standard deviation 2.5 inches. Half of all young men are shorter than

a. 65.5 inches

b. 68 inches ← Since its normal, half (or 50%) of the population is less than the mean.

c. 70.5 inches

d. can't tell, because the median height is not given

---

9. A set of data has a median that is much larger than the mean. Which of the following statements is most consistent with this information

a. a histogram of the data is symmetric ← will not be symmetric otherwise mean = median

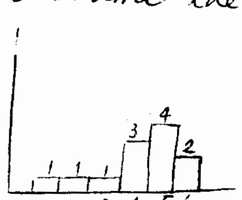
b. a histogram of the data is skewed left ← median > mean

c. a histogram of the data is skewed right ← what do you think the comparison of the median and the mean is?

d. Not enough information to answer the question.

Assume the data set is 1, 2, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6

Obviously it's skewed to the left.



$$\left.\begin{array}{l} \text{median} = \frac{4+5}{2} = 4.5 \\ \text{mean} = \frac{50}{12} = 4.17 \end{array}\right\} \text{so median} > \text{mean}!$$

$\bar{X}$ $\quad$ n

10 - 11. The mean federal income tax paid last year by 100 persons selected from a city was \$4000. It is known that the standard deviation of the federal income tax paid by all people in the city is \$200. USE WHAT YOU KNOW ABOUT THE 68-95-99.7 RULE TO ANSWER THIS QUESTION.

$z=2$ $\quad$ σ

a. Construct a 95% confidence interval for the mean federal income tax paid last year by all persons living in this city

$$95\%\ CI = \left(\bar{X} - z\cdot\frac{\sigma}{\sqrt{n}},\ \bar{X} + z\cdot\frac{\sigma}{\sqrt{n}}\right)$$

$$4000 - 2\cdot\frac{200}{\sqrt{100}} \qquad 4000 + 2\cdot\frac{200}{\sqrt{100}}$$

$$= 4000 - 40 \qquad = 4000 + 40$$

$$= 3960 \qquad = 4040$$

$$\therefore\ 95\%\ CI = (\$3{,}960,\ \$4{,}040)$$

b. To make the margin of error smaller, should the sample size be increased or decreased?

The margin of error is $z\cdot\frac{\sigma}{\sqrt{n}}$. To make it smaller, increase n since dividing a larger n, will make a quantity smaller.

12. A Math Placement Exam is given to test takers across the country. A sample of 100 test takers has a mean of 70 on this exam. Such scores are known to have a standard deviation of 40. Use the 68-95-99.7 rule-of-thumb to find the 95% confidence interval of the true mean score on this placement test.

n $\quad$ $\bar{X}$ $\quad$ σ $\quad$ $z=2$

$$\bar{X} \pm z\cdot\frac{\sigma}{\sqrt{n}}$$

$$= 70 \pm 2\cdot\frac{40}{\sqrt{100}}$$

$$= 70 \pm 2\cdot(4)$$

A. 70 ± 40 or from 30 to 110

B. 70 ± 2(40) or from -10 to 150

C. 70 ± 2(4) or from 62 to 78

D. 70 ± 1.96(4) or from 62.16 to 77.84 ← *This would be the answer if they didn't mention the 65-95-99.7 rule.*

E. 70 ± 4 or from 66 to 74

13. In a classroom full of students, the mean score on an exam is 69 while the median score is 72. Use this information to answer the following questions; circle True or False for each question below.

a. True or False: The graph of the distribution of the exam scores is skewed to the left. (See #9)

b. True or False: The variance of the scores is zero.
The only way the variance (and hence the standard deviation) is zero if and only if all the numbers in the data set are the same. However, since mean is 69 ≠ median is 72, therefore, we can't have all the numbers are the same.

14. In the very same classroom in (13) above, a new student takes the exam and scores a 69. Use this information to answer the following questions; circle True or False for each question below.

a. True or False: The new standard deviation will be larger than the original. The new S.D. will be actually smaller.

b. True or False: The new mean is larger than 69.
The new mean will stay the same.

The following is the effect of inserting a new number into an existing data set.

If the	old mean and	old SD, and	insert,	then new mean and	new SD
①	70	10	100	>70	>10
②	70	10	75	>70	<10
③	70	10	70	=70	<10
④	70	10	65	<70	<10
⑤	70	10	40	<70	>10

- In general, inserting a number higher/lower/equal to the mean of the existing data set will make the new mean higher/lower/equal.
- In general, inserting a number close to/far away from the mean of the existing data set will make the new standard deviation smaller/bigger.

15 - 18. In a 10 year study of the biological and demographic characteristics of the liefish, data were collected on the mean length of specimens collected each year for liefish whose ages ranged from 2 though 8 years.
Find the least squares prediction equation relating the mean length y of the liefish to its age x . USE THE TABLE BELOW TO WRITE THE EQUATION:

Dependent variable is: **Mean Length (mm)**
No Selector
R squared = 97.4% R squared (adjusted) = 96.9%
s = 4.943 with 7 - 2 = 5 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	4577.29	1	4577.29	187
Residual	122.143	5	24.4286	

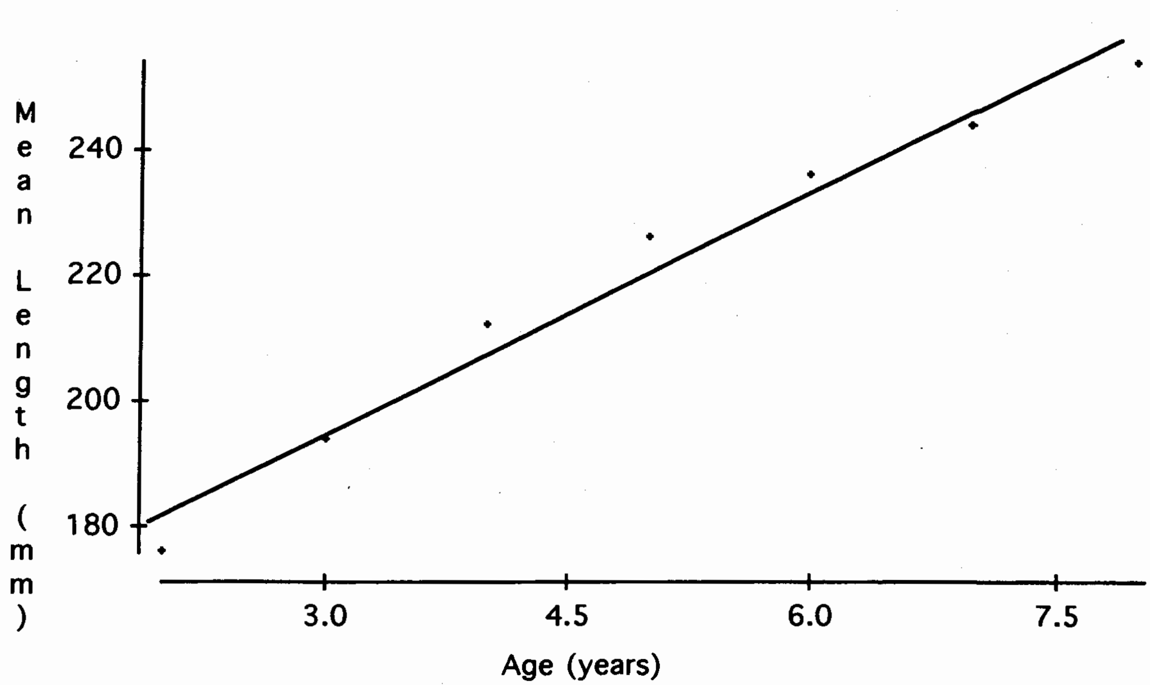
Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	156.357	5.030	31.1	≤ 0.0001
Age (years)	12.7857	0.9340	13.7	≤ 0.0001

a. EQUATION: $\hat{y} = 12.7857x + 156.357$

b. About how long would a liefish be when it is 10 years old? Show your work.

$\hat{y} = 12.7857(10) + 156.357 = 284.214$ cm

Note: we use the linear regression equation since there is a signifance correlation between x (age) and y (length) and also the number 10 is still close to 8, the maximum age recorded.



c. We have plotted the data points and added a least squares line to the scatterplot on the previous page. Use the regression table to determine which of the following is closest to the correlation of age and length. Circle your answer.

1.0   0.8   0.6   0.4   0.2

Since R-squared $= 97.4\% = .974 \Rightarrow r = \sqrt{.974} = .987 \approx 1$

d. What value on the regression table did you use to find the answer to c.

The value of R-squared.

19 - 25. In the table below, some of the entries have been erased. Replace the missing numbers and use the table to answer the questions below.

Survey Answer

Class	YES	NO	NO ANSWER	TOTAL
SENIORS	20	10	0	30
JUNIORS	19	20	1	40
SOPHOMORES	21	20	0	41
Total	60	50	1	111

You pick one student at random. What is the probability that:

he is a SENIOR? $30/11$

he said NO and he is a SENIOR? $\frac{50}{111} + \frac{30}{111} - \frac{10}{111} = \frac{70}{111}$

he said NO or he is a SENIOR? $\frac{10}{111}$

he is a SENIOR or a JUNIOR? $\frac{30}{111} + \frac{40}{111} - 0 = \frac{70}{111}$

nobody is a senior and junior at the same time so P(Senior and Junior) = 0

he said NO given that he is a SENIOR?

$\frac{10}{30}$ ← 10 seniors said no; ← total number of seniors

he is a SENIOR, given that he said NO?

$\frac{10}{50}$ ← of the 50 people said no, 10 are seniors; ← 50 people said no.

he is a SENIOR, given that he said NO or he is a JUNIOR, given that he said NO?

$\frac{10}{50} + \frac{20}{50} = \frac{30}{50}$

Extra Problems

1. A Math Placement Exam is given to test takers across the country. A sample of 100 test takers has a mean of 70 on this exam. Such scores are known to have a standard deviation of 20. Use the 68-95-99.7 rule-of-thumb to find the 95% confidence interval of the true mean score on this placement test.

A. 70 ± 20 or from 50 to 90

B. $70 \pm 2(20)$ or from 30 to 110

C. $70 \pm 2(2)$ or from 66 to 74

almost same as #12

D. $70 \pm 1.96(2)$ or from 66.08 to 73.92

E. 70 ± 2 or from 68 to 72

2. The administration claims that the average height in this classroom is 5 feet 5 inches. A random sample of 36 students shows a mean height of 5 feet 7 inches with a standard deviation of 6 inches. Is the true height greater than 5 feet 5 inches?

Which of the following null and alternative hypotheses would you test in this situation?

A. Ho: μ = 5 feet 5 inches
Ha: μ > 5 feet 5 inches

B. Ho: μ = 5 feet 5 inches
Ha: μ = 5 feet 7 inches

C. Ho: μ = 5 feet 5 inches
Ha: $\mu \neq$ 5 feet 5 inches

D. Ho: μ = 5 feet 7 inches
Ha: μ = 5 feet 5 inches

E. Ho: μ = 5 feet 5 inches
Ha: μ < 5 feet 5 inches

3. The administration claims that the average height in this classroom is 5 feet 5 inches. A random sample of 36 students shows a mean height of 5 feet 7 inches with a standard deviation of 6 inches. Is the true height different from 5 feet 5 inches?

Which of the following null and alternative hypotheses would you test in this situation?

A. Ho: μ = 5 feet 5 inches
Ha: $\mu >$ 5 feet 5 inches

B. Ho: μ = 5 feet 5 inches
Ha: μ = 5 feet 7 inches

C. Ho: μ = 5 feet 5 inches
Ha: $\mu \neq$ 5 feet 5 inches

D. Ho: μ = 5 feet 7 inches
Ha: μ = 5 feet 5 inches

E. Ho: μ = 5 feet 5 inches
Ha: $\mu <$ 5 feet 5 inches

---

4. What type of test would you use for the following situation?

Ten students take both a pretest and post-test. Compare their before and after scores.

A. z-test of individual means

B. t-test of individual means

C. pooled t-Test of $\mu_1 - \mu_2 = 0$

D. 2-sample t-Test of $\mu_1 - \mu_2 = 0$

E. paired t-Test of $\mu_1 - \mu_2 = 0$

Since we want to see how the post-test scores individually compared to the pretest scores.

5. What type of test would you use for the following situation?

25 female students take a strength test and 35 male students take a strength test. Compare the male and female strength tests. Assume the variances are equal.

A. z-test of individual means

B. t-test of individual means

C. pooled t-Test of $\mu_1 - \mu_2 = 0$ ← Since they mention: "Assume the variances are equal"

D. 2-sample t-Test of $\mu_1 - \mu_2 = 0$ ← This would be the choice if they didn't mention the variances are equal

E. paired t-Test of $\mu_1 - \mu_2 = 0$

STATISTICS 113 SPRING 2006
LECTURE 12 Profs. Clarkson and Williams

GROUP COLOR AND LETTER ____________________

Name______________________________

Name______________________________

Name______________________________

Name______________________________

Name______________________________

A regression predicting temperature from CO_2 level produces the following output table (in part)

Dependent variable is: Temperature → dependent variable

R-Squared = 36%

Variable	Coefficient
Constant	19.3066
CO_2 Level	0.004

→ explanatory variable (or predictor): a variable which is used to explain or predict changes in the values of another variable. Here, we use CO_2 level to explain the changes in temperature

a. What is the correlation between CO_2 level and temperature?

$\sqrt{.36} = .6$

b. What is the name of the explanatory variable?

CO_2 level

c. What is the number value of the y-intercept?

19.3066

d. What is the equation of the relationship between temperature and CO_2 level?

$\hat{y} = .004x + 19.3066$

e. CO_2 levels will likely reach 1000 parts per million this year. Using that information, what mean temperature does the regression predict?

$\hat{y} = .004(1000) + 19.3066 = 23.3066$