

1. A poll on women's issues interviewed 1,025 women and 472 men randomly selected from the United States. The poll found that 47% of the women said they do not get enough time for themselves.

(a) Construct an exact 90% confidence interval for the percentage of women who say they do not get enough time for themselves.

$$.47 \pm 1.645 \left(\sqrt{\frac{(.47)(.53)}{1025}} \right)$$

$$.47 \pm .0256 \text{ or } (.4444 \text{ to } .4956)$$

(b) Your friend is taking Statistic 11 next year and unfortunately they are enrolled at 8am with Professor Lew and they ask you for help all of the time. Explain to your friend why we can't just say that 47% of all adult women in the U.S. do not get enough time for themselves.

You could say that 47% is the percentage in the sample, but what you would really like is the population percentage - which we do not have and we cannot be completely certain that the sample percentage is exactly equal to the population percentage because of chance error.

As long as we are working with samples and do not have the population, we are unable to give exact figures.

2. The following comes from a recent article in *The Wall Street Journal* :

Generation X-ers Aren't Relying On the Survival of Social Security

BY JOHN SIMONS

According to the most recent *Wall Street Journal*/NBC News poll, only 39% of X-ers believe that Social Security will still be able to provide benefits when they retire. That compares to recent surveys of all Americans which show that 45% think so.

Assume that 45% figure is a stable, long-run, historical fact about American beliefs about Social Security benefits. Also assume that the survey of Generation X-ers had 121 respondents.

- (6.2/6.3) A. Test the hypothesis that the belief that Social Security will still be able to provide retirement benefits has decreased over time. State the null and the alternative, perform a test, and state a p-value. Please use a 5% level of significance as your decision rule. On the basis of your test results, do you think that Generation X-ers are like other Americans in their beliefs about social security or are they different?

$$H_0: p = .45 \quad Z = \frac{.39 - .45}{\sqrt{\frac{(.45)(.55)}{121}}} = -1.33$$
$$H_a: p < .45$$

$$P(Z \leq -1.33) \rightarrow$$

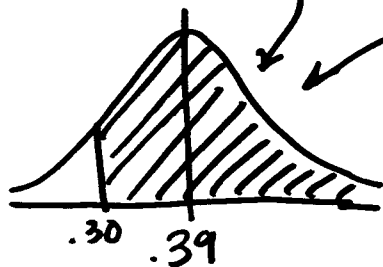
p-value is $\approx .092$ or 9.2%

9.2% > 5% do not reject null

the sample of Gen Xers do not appear to differ from the population of all Americans in their beliefs

- (5.1) B. Suppose in a few years the *Wall Street Journal* decided to replicate this study (i.e. draw a new sample) on Generation Y-ers (that's you all, I think...). Let's assume that the 39% figure is now the stable, long run fact about belief in Social Security benefits by Americans.

What is the chance that a sample of 64 will have at least 30% of the surveyed Generation Y-ers believing in Social Security?



$$Z = \frac{.30 - .39}{\sqrt{\frac{(.39)(.61)}{64}}} = -1.48$$

$$P(Z \geq -1.48) = 1 - .069 = .9306$$

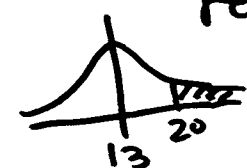
$\approx 93.1\%$

3. A simple random sample of 100 stocks was drawn from the entire market. The average return was 13%, and the SD was 6%; furthermore, the distribution of percentage returns in the sample was close to normally distributed.

Based on these data, is possible to construct a 95% confidence interval for the percentage of stocks in the market as a whole that had percentage returns greater than 20%?

Please answer yes or no. If you answer yes, please construct a confidence interval and please explain why this is possible to do. If you answer no, please explain why it is inappropriate to construct a confidence interval given this information.

(1.3) YES. The distribution is close to normal, so we can use normal calculations (i.e. z) to figure out what proportion of ^{all} stocks have % returns > 20%.



$$z = (20 - 13) / 6 = +1.17 \quad P(z \geq 1.17) = .1210$$

(6.1) → for 95% confidence then

$$.1210 \pm 1.96 \left(\sqrt{\frac{(.1210)(.8790)}{100}} \right) \Rightarrow (.0571, .1849)$$

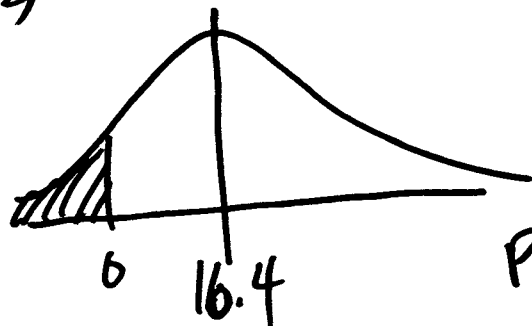
4. 1996 was a particularly good year for the stock market. For 1996 as a whole, the mean return on all common stocks on the NYSE (New York Stock Exchange) was $\mu = 16.4\%$. The standard deviation was about $\sigma = 36\%$. Assume the distribution of annual returns is roughly normal.

(5.2) (a) Suppose you selected 9 stocks at random from the NYSE stocks in 1996. What is the expected value (mean) and standard deviation of the returns of randomly chosen portfolios of 9 stocks?

$$\mu = 16.4$$

$$SD = \sigma / \sqrt{n} = \frac{36}{\sqrt{9}} = 12$$

(5.2) (b) What percentage of such portfolios of 9 stocks will lose money (i.e. have returns of zero or less)?



$$z = \frac{0 - 16.4}{12} = -1.37$$

$$P(z \leq -1.37) = .0853 \text{ or } 8.53\%$$

(1.2)

5. Here are two statistics on all persons who consider themselves investment bankers in 1997:

\$820,000 dollars per year

\$141,000 dollars per year

Which one of these numbers is the mean salary from investment banking and which one is the median salary from investment banking in 1997? Assume the samples were of good quality.

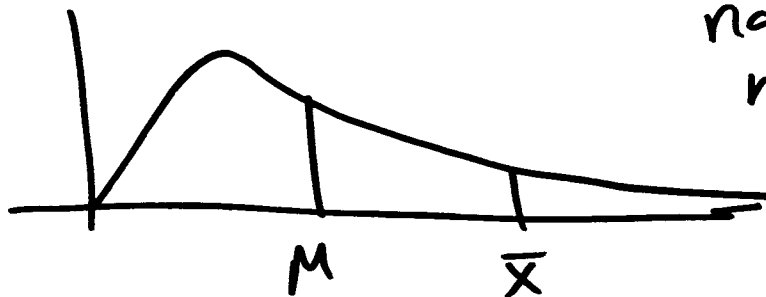
The mean is 820,000

The median is 141,000

Explain your choice in the space below. Be brief. This is not a long answer.

1) The minimum salary is 0 but the max can be very large

2) So the dist looks like



right skewed
mean > median

6. You got a job at an insurance company. The company advertises that it processes 90% of its claims on time, that is, within 5 working days of initial receipt. The average processing time of all claims is 2.7 days with a standard deviation of 0.6 days. You have been asked to select and audit a simple random sample (SRS) of 42 of the tens of thousands of claims received in the past month. The audit reveals that 37 of the 42 claims were processed within 5 working days with an average of 2.9 days and a standard deviation of 1.1 days.

(5.1) (a) What are the ^{expected} mean and standard deviation for the number of claims processed on time?

$$\text{mean} = \sim 37.8 = np = (42)(.90)$$

$$SD = \sqrt{(42)(.90)(.10)} = 1.94 \text{ claims}$$

(6.2) (b) An angry client thinks the processing time of 2.7 days is a lie and in reality, the company is slower than it advertises. Please test the hypothesis that the true processing time is 2.7 days against the client's alternative. Is there sufficient evidence to reject the null at $\alpha = .05$?

$$H_0 : \mu = 2.7$$

$$H_a : \mu > 2.7$$

$$z = \frac{2.9 - 2.7}{0.6 / \sqrt{42}} = +2.16$$

$$P(Z \geq 2.16) = .0154$$

$$.0154 < .05$$

Yes. reject the null, company appears to be slower than advertised.

7. Here is some Stata output: (Ch. 2.1-2.4)

. regress totgross openday

Source	SS	df	MS	Number of obs	=	907
Model	4.9206e+17	1	4.9206e+17	F(1, 905)	=	278.78
Residual	1.5974e+18	905	1.7651e+15	Prob > F	=	0.0000
				R-squared	=	0.2355
				Adj R-squared	=	0.2347
				Root MSE	=	4.2e+07
Total	2.0895e+18	906	2.3062e+15			

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
openday	23219.24	1390.652	16.697	0.000	20489.96 25948.51
_cons	714478.1	2426230	0.294	0.768	-4047214 5476170

totgross is Total gross receipts for a movie

openday was the number of screens the movie was being shown on its first day of viewing

a. Write out the regression equation estimated by Stata. Identify the slope and intercept clearly.

$$\text{totgross} = 714478.1 + 23219.24 (\text{openday})$$

(y) (intercept) (slope) x

b. What is the interpretation of the coefficient for the variable openday? Please state it in plain English and include how it relates to totgross.

for each additional screen showing the movie on its opening day, the total gross receipts were expected to increase by \$23,219

(continued from above)

- c. What is the correlation for totgross and openday? Please interpret its value in terms of both strength and direction.

$$\sqrt{.2355} = +.485 \approx .49$$

- d. A movie titled "How the Grinch Stole Christmas" opened on 3,127 screens across the United States on November 17, 2000. From this information, can you give me an estimate of its total gross receipts? Answer yes or no and then give me a numerical solution.

YES.

$$\boxed{73,321,041.58} = 714478.1 + 23219.24(3127)$$

- e. Suppose I told you that a movie titled "Miss Congeniality" had total gross receipts of \$103,271,534. From this information, can you tell me how many screens it was being shown on its opening day of December 22, 2000? Answer yes or no. If you answer yes, please give a numerical solution and show your work. If you answer no, please explain why this is not possible.

No. Screens is NOT the outcome variable, it is the predictor.

The equation + information above won't allow you to do this.

If you could recalculate the equation such that

$$\text{openday} = a + b(\text{totgross})$$

then you could answer part e.

8.

High Bias and High Variance are both considered undesirable features of certain sample statistics (such as a sample mean for example). You are working with a team on a marketing study, a sample of size 100 is drawn. One of the variables you are interested in is the average time spent on the internet on any day. You plan to construct confidence intervals and perform some unspecified hypothesis tests. Studies always have problems, and today you have your choice: High Bias or High Variance. Which one would you rather deal with and why?

Ch 3.4

I would rather deal w/ high variance because I could at least believe my estimate (even if it is not too useful) and perhaps there is a chance of salvaging the information if I can grab another sample and add to this one.

With bias - we don't know where the "truth" is (unless you know the exact nature of the bias)
Adding more ~~people~~^{n's} to the sample will not help eliminate bias.

I can lower variability - consider

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \sim \frac{\sigma}{\sqrt{n}}$$

but bias? Not easily addressed
as

4.4

+

5.2

9. You work for an investment bank and you are evaluating two companies A and B. Both of them are in biotechnology and each company has invested a significant amount of money in a number of projects.

Here is some information on the two companies:

$$(10 \times 100,000) = 1,000,000 \text{ spent}$$

Company A currently has 10 projects under development, each project cost the company \$100,000. Suppose when completed, each project has the following distribution: a 50% chance that the project will be worth \$500,000, a 20% chance that it will be worth \$100,000 otherwise, it will be worth nothing.

$$(20 \times 200,000) = 4,000,000 \text{ spent}$$

Company B currently has 20 projects under development, each project cost the company \$200,000. Suppose each project when completed has the following distribution: a 60% chance that the project will be worth \$400,000 otherwise, it will be worth nothing.

Assume the projects are independent and when completed and with the price of the initial investment factored in, the total of the projects will represent the total net worth of each company.

(a) What are the expected net worths for Company A and Company B?

$$\begin{aligned} \mu_A &= (500K \times .50) + (100K \times .20) + (0 \times .30) = 270K & b \cdot \mu_A &= 10 \times 270K = 2,700,000 \\ \mu_B &= (400K \times .60) + (0 \times .40) = 240K & b \cdot \mu_B &= 20 \times 240K = 4,800,000 \\ \text{net worth}_A &= 2,700,000 - 1,000,000 = 1,700,000 \\ \text{net worth}_B &= 4,800,000 - 4,000,000 = 800,000 \end{aligned}$$

(b) What are the Standard Deviations for Company A and for Company B?

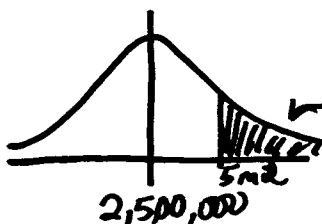
$$\begin{aligned} \sigma_A^2 &= (500K - 270K)^2(.50) + (100K - 270K)^2(.20) + (0 - 270K)^2(.30) = 54,100,000,000 \\ \sigma_B^2 &= (400K - 240K)^2(.60) + (0 - 240K)^2(.40) = 38,399,990,000 \\ b \sigma_A &= \sqrt{(10)^2(54,100,000,000)} = 2,325,940.67 & b \sigma_B &= \sqrt{(20)^2(38,399,990,000)} = 3,919,183.51 \end{aligned}$$

(c) Your bank is trying to determine whether a merger of the two companies is worth the time and expense. They have called you in to make the call, here is the information you need: the bank thinks the merger should occur if there is a greater than 25% chance that the combined net worth of the companies exceeds \$5,000,000. Answer "merge" or "do not merge" and show the work that lead you to your decision.

$$\text{expected (mean) combined net worth} = 2,500,000 = 1,700,000 + 800,000$$

What is the chance of exceeding 5 mil?

$$z = \frac{5,000,000 - 2,500,000}{\sqrt{(10)^2(54,100,000,000) + (20)^2(38,399,990,000)}} = \frac{+2,500,000}{4,557,411.55} = .55$$



$$P(z \geq .55) = 1 - .7088 = .2912$$

$$.2912 > .25$$

MERGE

10. Over his career, a basketball player has made 1,210 free throws and missed 214. The basketball player is about to shoot the ball again, what is his estimated probability of making a free throw (i.e. getting the ball into the basket)?

4.2

$$\frac{1210}{(1210+214)} \approx .85$$

the best estimate will be his long-run proportion (the "parameter")

11. A manufacturer of "zip" floppies advertises that 95 out of 100 floppies will have no problems. Suppose you buy a pack of 16 from a discounter and find that one fails to work. If the manufacturer's claim is true, what is the probability that one or more floppies fail in a pack of 16?

4.2

$$\frac{95}{100} = .95 \quad \text{or } 95\% \text{ work}$$

$$P(1 \text{ or more fail}) = 1 - P(\text{no failures})$$

$$.5598 = 1 - (.95)^{16}$$

$$\approx .56 \text{ or } 56\%$$