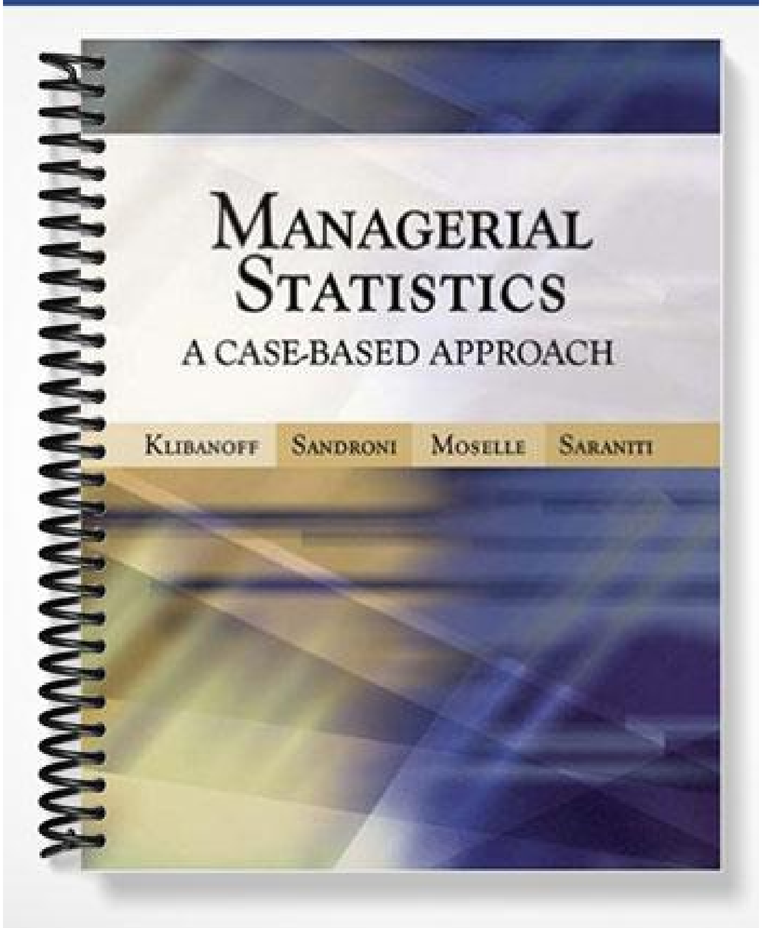
SOLUTIONS MANUAL



CHAPTER 2: ANSWER KEY

CASE EXERCISES

1: The Gender Gap

What can we say about the "plus or minus 4.5 percentage points"? Well, remember that providing a confidence interval simply means that rather than giving a single number (e.g., saying that "30% of likely voters favor Cruz Bustamante"), we are providing an estimate plus a margin of error (30% plus or minus 4.5%") to take into account the possibility of sampling error. So for example, we can calculate a 95% CI for the estimate that 30% of likely voters prefer Bustamante. The estimated standard deviation of the sample-proportion estimator $\overline{p}$ is

$$s_{\overline{p}} = \sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = \sqrt{\frac{(.30)(1-.30)}{505}} \approx .02$$

and the formula for a 95% CI is $\overline{p} \pm z_{.025} s_{\overline{p}}$, so the margin of error here is

$\pm z_{.025} s_{\overline{p}} = \pm(1.96)(.02) \approx 4\%$. In other words, we can be 95% confident that this figure of 30% is accurate to within plus or minus four percentage points (**not** 4.5 percentage points). The largest possible standard deviation will occur with a sample proportion of 50% when the standard error of the proportion equals

$s_{\overline{p}} = \sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = \sqrt{\frac{(.50)(1-.50)}{505}} \approx .0223$ yielding a margin of error of (1.96)(0.0223) = 4.4%

b. Subgroups will always have larger ranges since they have fewer data points than the entire group. In our case, the less accurate ranges are those which compare men and women which splitting the sample into two.

For Schwarzenegger, the standard error of the proportion of women who favor his candidacy is given by $s_{\overline{p}} = \sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = \sqrt{\frac{(.26)(1-.26)}{252.5}} \approx .028$ giving a margin of error of (1.96)(0.028) = 5.5%. For men the answer is a very similar 5.6%.

The Field Poll could have given different margins of error for the different estimates, but for most people reading the poll this would have been uninteresting or possibly confusing. The lesson is that once you have done your statistics you need to present your results in a concise and economical way, bearing in mind your target audience.

c. Arnold Schwarzenegger got favorable ratings of 29% from men in the sample, 26% from women. Is that difference statistically significant? During the campaign that question was certainly worth asking, since both sides thought a lot about the "gender gap" in formulating their strategies. Let's do the math: we have a difference of .03, and a standard deviation of

$$s_{\bar{p}_1-\bar{p}_2} = \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} = \sqrt{\frac{(.29)(1-.29)}{252.5} + \frac{(.26)(1-.26)}{252.5}} \approx .04$$

so the difference is less than one standard deviations, which we know is not that large. So informally, we know that there is not strong evidence that men favored Arnold more than women did (at least at the time of this poll.) We can do this more mechanically by writing down the hypothesis test and calculating the test statistic:

H_0: $p_1-p_2=0$

H_a: $p_1-p_2\neq 0$

The test statistic is

$$z = \frac{\bar{p}_1 - \bar{p}_2 - (p_1 - p_2)}{s_{\bar{p}_1-\bar{p}_2}} = \frac{.29 - .26}{.04} = 0.75$$

Excel gives us a p-value of 0.453 (using "=2*(1-NORMSDIST(0.75))") So we should accept the null hypothesis and conclude that there is not really any difference.

d. The calculation for Tom McClintock is identical, except for the numbers. The standard deviation is now

$$s_{\bar{p}_1-\bar{p}_2} = \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} = \sqrt{\frac{(.16)(1-.16)}{252.5} + \frac{(.10)(1-.10)}{252.5}} \approx .03$$

(lower than before… the further the sample proportions are from 0.5 the smaller this standard error will be) and the test statistic is

$$z = \frac{\bar{p}_1 - \bar{p}_2 - (p_1 - p_2)}{s_{\bar{p}_1-\bar{p}_2}} = \frac{.16 - .10}{.03} = 2.0$$

giving a p-value of 0.046 (again using Excel to compute "=2*(1-NORMSDIST(2.0))".) So we should reject the null hypothesis. Unlike the case of Arnold Schwarzenegger, there does seem to be a gender gap in the feelings of likely voters towards Tom McClintock.

e. For Gray Davis we use the same technique as before for all three cases: actual sample size, hypothetical 1000 surveyed, and hypothetical 10,000 surveyed.

	Actual	1000	10,000
Standard Error	0.434	0.031	0.0098
Test Statistic	-0.69	-0.97	-3.07
p-value	0.49	0.33	0.0021

In the *actual* case, the p-value says that, assuming there is no real difference between men and women in their support for Gray Davis , we can expect to see a difference of 3% or greater in samples of this size about half the time (in fact, 49% of the time), so that we certainly should not interpret the 3% difference in our sample as telling us that there is a real difference in the population. Increasing the sample size and recalculating the standard deviation and test statistic, we get z = -0.97 and a p-value of 0.33, so we would have around 67% confidence that there really is a difference (so still could not reject the null at 10% significance). Increasing the sample size to 10,000 and recalculating gives z = -3.07 and a p-value of 0.0021, so we could be absolutely certain that the difference was real.

You can draw a few lessons from this exercise. First, the bigger the sample size the more accurate the estimate, and hence the easier to draw conclusions from it. Second, for small differences to be statistically significant you need a very large sample. Third, statistically significant doesn't mean it's important: *so what* if there was a 3% gender gap in Gray Davis' support?

2. The January Effect

First, for a January effect on small cap companies: you should have got 0.004803 as the p-value. Since the p-value is very small (just half a percent), we can reject the null hypothesis (that returns are the same in January as in other months) at a very high level of significance. There really does seem to be a January effect in small cap stocks.

For the S&P500, we get a p-value of 0.475286, which is very high, so we certainly cannot reject the null. There is not strong evidence for a January effect in the S&P500 returns. Finally, for election years we get a p-value of 0.091525, so there is reasonably strong evidence: we can accept the alternative hypothesis, that t-bill returns are different in election years, at a 10% level of significance, but not much higher.

3: Fast Food Nation

a. The first step is calculating the standard deviation.

$$s_{\bar{p}_m - \bar{p}_w} = \sqrt{\frac{.50 \cdot (1-.50)}{770} + \frac{.62 \cdot (1-0.62)}{236}} = 0.03637$$

This then yields a test statistic of

$$z = \frac{0-(0.50-0.62\ )}{0.03637} = -3.23$$

For a one tailed test, this test statistic gives us a p-value of 0.0005 which tells us to reject the null hypothesis and conclude that people who think fast food is good actually do eat it more often.

b. The company believes that at least half of the target group will eat fast food more often if they offer healthier menu options. The Gallup data supports the opposite position which is that *less* than half of the members of this group would eat fast food more often; however, we don't know if their evidence is strong enough to sway the company's opinion.

The sample proportion is 84/204 = 0.41. Using this we determine that the sample standard deviation is $\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\frac{0.41(1-0.41)}{204}} = 0.0344$

The test statistic is equal to $z = \frac{0.41-0.50}{0.0344} = -2.56$ giving us a p-value of 0.0052. Since this p-value is very small, we can reject the null and conclude that less than half of the infrequent diners would eat fast food more often. The company should change its mind.

4: Pro Bowling for Dollars

This is a straightforward computation using the standard techniques from this chapter. Set up the hypothesis test as follows:

H_0: $\mu \geq 10.1$

H_a: $\mu < 10.1$

And then calculate the test statistic as:

$$t = \frac{8.6 - 10.1}{5.7 / \sqrt{260}} = -4.24$$

This gives us a p-value of 0.0000156 (use the command in Excel "=tdist(4.24,259,1)" to get this p-value) which clearly indicates that the average Pro Bowl visitor spends less than 10.1 days on their visit to Hawaii.

PROBLEMS

1. a & b. First let's set up the hypothesis test. We want to show that the true proportion is greater than 0.25 and so we make that the alternative hypothesis.

H_0: $p_s \leq 0.25$

H_a: $p_s > 0.25$

Analyzing the data gives us the following important facts and computations:

	Summer
Sample Proportion	0.255474
Sample Standard Deviation	0.436128
Standard Error	0.037261
Test Statistic	0.146922
p-value	0.441705

Thus, we accept the null and conclude that the proportion of movies released in the summer is not greater than one fourth.

c & d. Again we'll do the test first. We want to show that the true proportion is not equal to 0.10 and so we make that the alternative hypothesis.

H_0: $p_s = 0.10$

H_a: $p_s \neq 0.10$

Analyzing the data gives us the following important facts and computations:

	Holiday
Sample Proportion	0.131387
Sample Standard Deviation	0.337823
Standard Error	0.028862
Test Statistic	1.087474
p-value	0.278751

The relatively large p-value tells us that we should again accept the null hypothesis and conclude that the conventional wisdom is true.

2. a & b.

	MPAA Rating			
	R	PG13	PG	G
Mean	33.04655	57.66502	42.10632	71.3295
Standard Deviation	39.10435	50.59916	37.97772	65.36909

Organizing the data by rating and using Excel's AVERAGE and STDEV functions make these parts very straightforward.

c, d & e. For each pair of categories the null is that the true difference between their means is zero and the alternative is that the difference is not zero. Let's do one of the tests using the formula from the chapter:

H_0: $\mu_R - \mu_{PG13} = 0$

H_a: $\mu_R - \mu_{PG13} \neq 0$

The test statistic is given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\left(s_1^{\,2}/n_1\right) + \left(s_2^{\,2}/n_2\right)}} = \frac{(33.05 - 57.67) - 0}{\sqrt{\left(39.1^2/67\right) + \left(50.60^2/42\right)}} = -2.69$$

which gives us a p-value of 0.0083 (the p-value comes from Excel using a t distribution with n_1+n_2-2 (giving 67+42-2 = 107) degrees of freedom.) We should reject the null and conclude that there is a difference in the gross of films with R and PG13 ratings.

f. The table below indicates the p-value for each test which were computed using Excel's TTEST function.

	R	PG 13	PG
G-13	0.008904		
PG	0.341504	0.172263	
G	0.213606	0.640903	0.335379

Notice that the p-value from this table is not exactly the same as the one we derived by hand. This is because we used an approximation for the degrees of freedom in the hand calculation. One last observation here is the importance of sample size. The difference between the sample means of R & G is bigger than the gap between R & PG13 but the p-value for that test is much larger. Even though the absolute difference is greater with the G movies, we are less sure about the significance since there are only a handful of G movies in the sample.

3. a & b.

	Genre			
	Drama	Comedy	Action	Horror
Mean	41.3122	38.94378	58.02131	30.0869
Standard Deviation	44.75745	47.8724	51.05734	13.3436

Organizing the data by rating and using Excel's AVERAGE and STDEV functions make these parts very straightforward.

c, d & e. For each pair of categories the null is that the true difference between their means is zero and the alternative is that the difference is not zero. Let's do one of the tests using the formula from the chapter:

H_0: $\mu_{Drama} - \mu_{Comedy} = 0$

H_a: $\mu_{Drama} - \mu_{Comedy} \neq 0$

The test statistic is given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\left(s_1{}^2/n_1\right) + \left(s_2{}^2/n_2\right)}} = \frac{(41.31 - 38.94) - 0}{\sqrt{\left(44.76^2/30\right) + \left(47.87^2/41\right)}} = 0.214$$

which gives us a p-value of 0.83130 (the p-value comes from Excel using a t distribution with n_1+n_2-2 (giving 30+41-2 = 69) degrees of freedom.) In this case we will accept the null hypothesis and conclude that there is no significant difference in the gross of films between these two genres.

f. The table below indicates the p-value for each test which were computed using Excel's TTEST function.

	Comedy	Action	Horror
Drama	0.831343	0.187343	0.229795
Comedy		0.119546	0.307404
Action			0.010735

Notice that the p-value from this table is not exactly the same as the one we derived by hand (if you go out to five or six decimal places.) This is because we used an approximation for the degrees of freedom in the hand calculation. The approximation is pretty good in this case since the two sample standard deviations are pretty similar. You'll see the difference a little more measurably in the cases where the standard deviations are further apart.

4. a. For this problem, the data is organized such that Kstat's Univariate Statistics function will get the answers quickly for us. You can also derive these with Excel directly using the AVERAGE and STDEV functions.

Univariate statistics				
	Honolulu	**Hawaii**	**Kauai**	**Maui**
mean	16513.71	11869.23	12711.61	13647.97
standard deviation	8550.452	5446.494	6471.645	6977.79
standard error of the mean	1535.707	978.2192	1162.342	1253.248

b & c. Each of the six tests will have the same format. The null is that there is no difference between the two means and the alternative is that the two means are not equal. For instance,

H_0: $\mu_{Honolulu} - \mu_{Maui} = 0$

H_a: $\mu_{Honolulu} - \mu_{Maui} \neq 0$

Doing this one test by hand is fairly straightforward. First we compute the test statistic as follows:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}} = \frac{(16514 - 13648) - 0}{\sqrt{(8550^2/31) + (6978^2/31)}} = 1.45$$

d. Using Excel's TDIST function for 31+31-2 = <u>60</u> degrees of freedom ("=TDIST(1.45,60,2)") we can determine the p-value to be 0.1534. The other value for the five remaining tests should be pretty close to the ones given in the table below for part e.

e. These p-values are directly from Excel's TTEST function:

	Hawaii	Kauai	Maui
Honolulu	0.013798	0.053321	0.153658
Hawaii		0.58136	0.267934
Kauai			0.58587

These should be close but probably not identical to the ones you derived in part d.

5a. Again Kstat's Univariate Statistic function gets us quickly to the information:

Univariate statistics		
	Weekday	**Saturday**

mean	337.835294	318.447059
standard deviation	58.8583544	60.9829223
standard error of the mean	6.38408488	6.61452662

b. & c. The hypothesis tests are both the same:

H_0: $\mu = 330$

H_a: $\mu \neq 330$

Where μ is the mean number of brochures taken at the banks.

For Weekdays the test statistic is (337.8-330)/6.38 = 1.23 giving a p-value of 0.22 and so we accept the null hypothesis and conclude that 330 people are taking the brochures. On Saturdays we compute the test statistic to be (318.4-330)/6.61 = -1.75 giving a p-value of 0.084 which again tells us to accept the null hypothesis and conclude that 330 people are taking the brochure on Saturdays.

d. We use a difference of means test for this part.

H_0: $\mu_{Weekdays} - \mu_{Saturday} = 0$

H_a: $\mu_{Weekdays} - \mu_{Saturday} \neq 0$

Excel's TTEST can do this problem in a hurry. The answer is: p = 0.0364 which tell us to reject the null hypothesis and conclude that there is a difference in the Weekday mean and the Saturday mean. This can be confusing since we concluded in parts b and c that both means were no different from 330.

The difference between a logical proof and a statistical proof should be clear with this problem. Each mean is shown to be the same as 330 but they are also shown to be different from one another. How can this be? Think about it a different way: we can't prove that either of the means is different from 330 but we can prove that they are not identical to each other. Just because we can't prove that something is true doesn't mean that it's false. It just means that we don't have strong enough evidence.

6. We want to conduct the same hypothesis test for each stock:

H_0: $\mu \leq 0$

H_a: $\mu > 0$

Because the data is the excess returns which equals the actual returns - t-bill rate, then the null hypothesis is equivalent to stating that:

actual returns - t-bill rate ≤ 0 or actual returns $\leq$ t-bill rate

If we have a low p-value then we should reject the null and conclude that the actual returns are greater than the t-bill rate. Kstat and Excel will give you the following results:

	EVANS	DAIRY	MACS	SIZZLER	WENDYS
Mean	0.470455	1.366667	1.1	0.195152	0.519621
standard deviation	7.18571	8.728106	6.232106	10.38666	10.01031
standard error of the mean	0.625436	0.759684	0.542435	0.904043	0.871286
test statistic	0.752203	1.798993	2.027892	0.215865	0.596384
p-value	0.226639	0.037161	0.0223	0.414714	0.275974
Outcome	Accept	Reject	Reject	Accept	Accept

While DAIRY has the highest average excess returns, MACS is the stock with the lowest p-value (although they are both pretty low.) The reason for this difference is because MACS has a smaller standard deviation than DAIRY. That makes us more certain about the fact that the average excess returns for MACS is larger than zero. Looking at the test statistic for both we see that while DAIRY has a higher numerator (1.37 vs. 1.10), MACS has a smaller denominator (0.54 vs. 0.76) which in this case results in the higher number and thus the lower p-value.

7a. Kstat gives us the relevant information using the Univariate Statistics function:

Univariate statistics

	Net Worth 2002	Net Worth 2001
mean	5762	6178
standard deviation	6620.228871	7345.246247
standard error of the mean	662.0228871	734.5246247

b. The best answer to this question is "yes, sort of." The mean net worth of the top 100 Americans was lower in 2002 than it was in 2001, but this fact could be misleading. The problem is that the wealthiest 100 Americans are not the same people form year to year. It's possible that some very wealthy people died and gave their money away to charities or split it up among 20 grandchildren. All we know is that the wealthiest 100 in 2002 are not as well off as the wealthiest 100 in 2001.

c. The univariate statistics give us a quick informal answer which is no. The drop in wealth is only about 400 million (oh, that's all) while the standard errors are both over 650 million. That difference is too small to be significant, but let's use Excel's TTEST function to be sure. For a one tailed test, the function gives us a p-value of 0.3372 which tells us that the drop in wealth is not significant.

8a. Here's the Univariate statistics from Kstat:

Univariate statistics

	Age 2002	Age 2001
mean	65.87	64.36
standard deviation	12.3342792	12.2454032
standard error of the mean	1.23342792	1.22454032

b. The mean age definitely increased from 2001 to 2002 by about 1.5 years. This signals a clear change in the make up of the membership of this group. Some of the younger members have been replaced by older wealthy people. If the entire group had been the same, the mean age would have increased by one year.

c. There are two ways to test the significance of this change just like the previous problems and the simplest is to use Excel's TTEST function. You should get a similar answer if you use the formulas from the text and generate a test statistic etc… TTEST gives us a p-value for a two tailed test as 0.386. You could reasonably have set this up as a one tailed test as well in which case you should get 0.193.