Solutions 7

Review, one sample t-test, independent two-sample t-test, binomial distribution, standard errors and one-sample proportions.

(1) Here we debunk a popular misconception about confidence intervals and practise obtaining p-values for one-sample statistics.

We are interested in the distribution of 6 week old calves. In Figure 1 the relative frequency histogram is given of the 44 calves and the output for constructing a 90% confidence interval for the mean weight of a 6 week old calf.

(i) By comparing the confidence interval to the distribution of weights, explain why the confidence interval is NOT an interval that contains 90% of calf weights and explain what information the confidence interval DOES convey.

From the plot we see that the confidence interval $[113.85, 120.87]$ contains less than 25% of the weights of the calves. This illustrates that a CI does not contain information about the distribution of weights. The sole purpose of the CI is to locate the mean weight of calves. We say that with 95% confidence the mean weight lies in the interval $[113.85, 120.87]$. From the histogram it appears that 95% of the weights lie between $[100, 140]$ pounds. As the sample size grows the 95% CI interval gets smaller, but the interval in which 95% of the weights lie will remain the same.

(ii) In previous years the mean weight of 6 week old calves was 112 pounds, there has been some speculation that the mean weight of calves has increased, after the shift from grass to corn feed in the calves diet. In Figure 1 we can see a histogram and Statcrunch output for 44 randomly selected corn fed calves. Based on this, in the following questions we see if there is evidence of an increase in weight for corn fed calves.

(a) State your null and alternative hypothesis of interest.

We are interested in see whether there is evidence to suggest that on corn feed became heavier. Thus we are testing $H_0 : \mu = 112$
against $H_A : \mu > 112$. The sample mean is $\bar{x} = 117.36$, below we will calculate how likely we are to observe a sample mean of 117.36 when the true population mean is 112.

(b) Using Figure 1 evaluate the t-transform associated with the above hypothesis. The t-transform is $t = \frac{117.36 - 112}{2.086} = 2.56$

(c) In Statcrunch go to Statistics $\rightarrow$ Calculator $\rightarrow$ T and calculate the p-value associated with the above t-statistic and test. As the alternative hypothesis is pointing to the right ($H_A : \mu > 112$) we need to calculate the area to the right of 2.56. We see from the Figure below that this is 0.7%.

![Figure 2:](image)

(d) The test is done at the 5% level, based on the above p-value is there evidence to suggest that corn feed may be increasing the weight of calves? As the p-value is 0.7% and this is less than 5%, this suggests that the alternative hypothesis (that the mean weight has increased) is true.

(2) Here we revise transformations of data.

In midterm II the range of grades was from 5 marks out of 15 to 13 marks out of 15. The average grade is 9.1 and the standard deviation is 1.9.

The professor wants to rescale the grades such that the highest grade is 15 out of 15. She has two options:

(A) Add two points to all the students scores. Eg. If a student scored 8 their new score will be 10.

(B) Multiply all the students scores by the factor (15/13). Eg. If a student scored 8 their new score will be $9.230769$.

Answer the following questions:

(i) What will the new highest and lowest grades be for both option (A) and option (B).

Option A: lowest grade = 7 and highest grade = 15

Option B: lowest grade = 5.769.. and highest grade = 15.
(ii) What will be the new average grade be for option (A) and option (B)?

Here we use the formulate if \( y = ax + b \), then \( \text{mean}(y) = a \times \text{mean}(x) + b \).

Option A: new mean = 11.1
Option B: new mean = 10.5

(iii) What will be the new standard deviation for option (A) and option (B)?

Here we use the formulate if \( y = ax + b \), then \( \text{sd}(y) = a \times \text{sd}(x) \)

Option A: new standard deviation = 1.9 Option B: new standard deviation = \((15/13) \times 1.9 = 2.19\)

(iv) Using the above information compare the spread of grades using option (A) and option (B).

We see that using option A the grades are simply shifted upwards, so the mean changes by the same shift but the spread of the data (as seen by the new range \([7,15]\) and standard deviation, \(s = 1.9\)) remains the same.

On the other hand using option B, where the grades are multiplied by a factor, leads to an increase in mean (which is less than option A), but the spread of the data has increased, as seen by the new range \([5.769,15]\) and new standard deviation (2.19).

Option B tends to penalise more lower scoring students.

(3) What is wrong with each of these statements?

(i) A researcher wants to test \( H_0 : \bar{x}_1 = \bar{x}_2 \) against \( H_A : \bar{x}_1 \neq \bar{x}_2 \).

The hypothesis are stated in terms of the sample means, which makes no sense. We are not interested in the samples (as these will vary, and it is very unlikely that two sample means will ever be the same). But we are interested in the population (which is fixed).

(ii) A study recorded the IS scores of 100 college students. The scores of 56 males in the study were compared with all 100 students in the study using the independent two-sample t-test.

Two samples, where one sample is a subset of another (as in the case described above), are definitely not independent. Therefore it is inappropriate to apply the independent two-sample t-test to such samples.

(iii) A two-sample t statistic gave a p-value of 0.94. From this we can reject the null hypothesis with 90% confidence.

If the p-value is 0.94, then it is extremely large, we would NOT be able to reject the null hypothesis at any meaningful significance level. Do not get signifiance level confused with confidence interval, they are very different.

In fact if the p-value is 0.94 it would be contained within a 90% confidence interval.
(iv) A researcher is interested in testing $H_0 : \mu_1 - \mu_2 = 0$ against $H_A : \mu_1 - \mu_2 < 0$. The test gave $t = 2.15$. Since the p-value for the two-sided alternative ($H_A : \mu_1 - \mu_2 \neq 0$) gave p-value equal to 3.6%, the researcher concludes that the p-value for the one-sided test is 1.8%.

The alternative hypothesis $H_A : \mu_1 - \mu_2 < 0$ is pointing to the left. This means we need to calculate the probability to the LEFT of 2.15 (which immediately means the p-value will be greater than 50%).

We can calculate the p-value by noting that the area to the RIGHT of 2.15 is 1.8%, therefore the area to the LEFT of 2.15 = $100 - 1.8\% = 98.2\%$, ie. the p-value = 98.2%.

(4) In this question we consider the meaning of a probability and apply the Binomial distribution.

Ricky’s friends think that Ricky has no clue about statistics. However, in the last multiple choice midterm Ricky scores 13 out of 15. In this question we will try to see how likely he could get this score by luck alone.

(i) Suppose that for each multiple choice question there are 5 different options (A, B, C, D, E). Assume that only one of these options is correct (no partial credit is given). What is probability that he will get a question correct by randomly choosing the answer?

It is one out of 5, which is 0.2.

(ii) State in terms of $\text{Bin}(n,p)$ the distribution of his scores if he were choosing the answer completely by random (ie. what is $n$ and $p$?).

$\text{Bin}(15, 0.2)$.

(iii) Using Statcrunch ($\text{Stat} \rightarrow \text{Calculators} \rightarrow \text{Binomial (use the =>) option when selecting the probability}$) calculate the probability of Ricky scoring 13 or over on a multiple choice just by random selecting the answers.

By using Statcrunch we see that it is the area to the right of (and including) 13 is extremely small ($5.7 \times 10^{-8}$). This is an extremely small probability.
(iv) Based on your answer in (iii), do you think Ricky is really that lucky or he is not the fool his friends think he is?

As the probability of Ricky scoring 13 out of 15 by just luck is extremely small, it seems unlikely that Ricky was just lucky and that actually he knew the material. If we pose this as a test we are testing $H_0 :$ Ricky chose the answers completely by random (ie the chance of choosing the correct answer is 0.2) against $H_A :$ Ricky was making an informed choice (ie. the probability of choosing the correct answer is greater than 0.2). We calculate the p-value under the null hypothesis (which is pointing to the right), as this is $5.7 \times 10^{-8}$%, which is extremely small (less than the usual 5% significance level) there is evidence to suggest that Ricky that the probability of Ricky getting an answer correct is greater than 0.2, thus he was making an informed choice when selecting the answer.

Ricky’s bestfriend Tricky is also taking the statistics class. In the same midterm he scored 5 out of 15. Using Statcrunch calculate the probability that Tricky scored 5 points or more given that he selected every answer completely by random. Based on this probability, which one of following seems the most likely interpretation of Trick’s performance in the midterm:

(A) Trick is clearly talented at statistics.

(B) Tricky must have guessed the answers to all of the questions.

(C) It is unclear whether Tricky was guessing or actually got those 5 questions correct because he knew their answer. However we do know that is quite plausible that someone can score 5 out of 15 just by randomly guessing the answer.

(D) If Ricky was guessing the answer we would expect him to get 3 questions correct. As he got more than this Ricky could not have been guessing.

Using Statcrunch the probability that Ricky got 5 marks or more by random is 16%. As this is a relatively large probability it means that he could have been guessing, though we do not know for sure. The correct answer is (C).

![Figure 4:](image-url)

If we were to pose this as a test than we are testing $H_0 :$ Tricky chose the answers completely by random (ie the chance of choosing the correct answer is 0.2) against $H_A :$ Tricky was making an informed choice (ie. the probability of choosing the correct answer is greater than 0.2). We calculate the p-value under the null hypothesis (which is pointing to the right), as this is $5.7 \times 10^{-8}$%, which is extremely small (less than the usual 5% significance level) there is evidence to suggest that Tricky was making an informed choice when selecting the answer.
answer is 0.2) against $H_A :\text{Tricky was making an informed choice (ie. the probability of choosing the correct answer is greater than 0.2)}$. The p-value is the probability to the right of 5 (as it is one-sided test pointing to the right), which is 16%. As this is larger than the 5% significance level there is no evidence to suggest that Ricky knew what he was doing, though it is impossible to tell whether he was guessing or actually knew the correct answers to those 5 questions.

(5) Our objective to compare the heights of French and German females. It is known that the standard deviation of French girls is 2.5 inches and the standard deviation of German girls is 3 inches. Using the formula for the standard error in the two-sample t-test $\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}$, calculate the standard errors in each of the following cases.

(A) Both samples have size 100.

The standard error for the difference in the means is

$$\sqrt{\frac{2.5^2}{100} + \frac{3^2}{100}}$$

(B) The number of French females 50 and the number of German girls is 150. The standard error for the difference in the means is

$$\sqrt{\frac{2.5^2}{50} + \frac{3^2}{150}}$$

(6) Which tests to do (either a one-sample test, a matched paired t-test or an independent two-sample t-test)?

(i) A meterologist wants to understand whether the mean temperature in College Station is the different to the mean temperature in Snook (which is about 20 miles away from College Station). In order to test this hypothesis she measures the temperature at 10am in College Station and Snook between April 6th and April 15th (10 readings from both towns).

**College Station and Snook share many meterological factors in common due to their close proximity. Therefore if we want to compare their temperatures it makes sense to consider their differences and do a MATCHED PAIRED T-TEST.**

(ii) We want to see whether children are taller than children 50 years ago. It was known that 50 years ago, the mean height of a child was 5 foot. This year a random sample of 500 children 10 year old children is drawn.

**There is only one sample in this situation. Therefore we need to do a one-sample t-test.**

(7) Our objective is to see if there is a difference in weights of eight week old calves given different treatments, in particular treatment B and treatment C. Two random samples (each of size eleven) of new born calves are given the different treatments. At 8 weeks, those on treatment B had a sample mean of 139.54 pounds and those on treatment
C had a sample mean of 144.45 pounds. This is difference of almost 5 pounds, to a layman this may suggest there is a difference in the treatments. However, by now you should realise that this difference could be due to sampling differences. Your objective is to analyse the data to see whether this difference is statistically significant (ie, the difference is too large and the standard error is too small so that probability that this difference is due to random is small). The output for the result is given in Figure 5, and the steps for answering this question are given below.

![Figure 5](image)

(i) Using the output evaluate the t-transform under the null hypothesis (that there is no difference).

We are testing $H_0 : \mu_C - \mu_B = 0$ against $H_A : \mu_C - \mu_B \neq 0$.

The t-transform is $t = \frac{-4.9}{6.74} = -0.72657$.

(ii) How many degrees of freedom do we use for the t-transform?

From the output it is clear that it is $df=19.97$.

(iii) Go to Statcrunch -> Stat -> Calculation -> $T$, and calculate the probability associated with the t-transform.

We use Statcrunch to calculate the area to the left of $-0.726$. This is $0.238$. Therefore the p-value is $2 \times 0.238 = 0.476$.

![Figure 6](image)

(iv) Based on your result in (iii), what is the p-value for the test and is there evidence to reject the null and suggest there is a difference in the mean weight of calves on treatment B and C?
As it is a two-sided test the p-value $= 2 \times 0.238 = 0.476 = 47.6\%$. As the p-value is a lot larger than any meaningful significance level (ie. 5% or 10%), then there is no evidence in the data to reject the null. There is no evidence to suggest that there is a difference between treatment B and treatment C.

(8) It is known that the proportion of people whose parents have had a university education is 30%.

There has been some speculation that there is an over-representation of this group at universities. In a random sample of 300 students, 110 of these students had parents who had a university education. Based on this sample we want to test if there is any evidence to test whether the proportion students with university educated parents is larger than the proportion 0.3.

(a) State the null and alternative hypothesis for the test of interest. It is a one-sided or a two-sided test?

$H_0 : p = 0.3$ against $H_A : p > 0.3$ (since we are seeing if there is any evidence of an over-representation of this group.)

(b) Now we calculate the chance of there being 110 students or more in the sample with parents who were university educated, when the overproportion of students with university educated parents is 0.3.

We do this in Statcrunch. Go to Statcrunch Stat $\rightarrow$ Calculators $\rightarrow$ Binomial. From here you can calculate the probability of there being at least 110 students in this samples whose parents went to university when the overall proportion of students whose parents went to university is 0.3.

What is this probability?

We see from Statcrunch that the probability is 0.00776.

![Figure 7:](image)

(c) Based on your answer (b) what is the p-value for the test and the result of the test.

As the alternative hypothesis is point to the right ($H_A : p > 0.3$), the p-value is the probability to the right (and including) of 110, which is 0.776%.
This probability is less than the 5% significance level, which suggests that it is unlikely that we would get a sample of 110 students out of 100 with parental university education when the true proportion at university was 0.3. Therefore there is evidence to suggest the proportion at university is higher than 0.3.

(d) A simpler way to do the test is to use the z-transform and normal approximation. Now we compare the results in (c) with the result using the z-transform.

Go to Statcrunch Stat $\to$ Proportions $\to$ One Sample $\to$ Summary. Place 110 in success and 300 in observations. Do the test (remember to correctly choose whether it is a one-sided test or a two-sided test).

What is the p-value for the test (using the z-transform) and what is the result? Using the Statcrunch output below the p-value is 0.59%. Therefore, as above using the normal approximation, there is evidence to reject the null hypothesis and determine that true proportion at university is higher than 0.3.

(e) Based on your results in (c) and (d), do you think there is much difference between using the exact method to obtain the p-value in (c) or the (normal) approximation given in (d)? Comparing the exact probability of 0.76% with the probability using the normal approximation of 0.59%, it seems that they are quite close. In particular do not give contradictory results in tests at the usual significance levels of 1%, 5% or 10%.