Course EPIB-621 - Data Analysis in the Health Sciences

Solutions to Assignment 1

1. In a one-sided frequentist test of the null hypothesis $H_0: \ \mu = 0$, versus $H_A: \ \mu > 0$, it is found that the p-value is p=0.0001. State whether each of the following statements are true or false, and explain why.

(a) After carrying out this experiment, there is only a one in ten thousand chance (0.0001) of being wrong if the conclusion is to reject the null hypothesis.

This statement is false. The p-value simply tells you that **if** *the null hypothesis is* **exactly** *true (i.e., to an infinite number of decimal places), then there is only a 1 in ten thousand chance of getting data as or more extreme than the data set observed. However, the p-value says nothing about the probability of being correct or incorrect in rejecting the null hypothesis, since it is calculated assuming the null hypothesis is true. Only a Bayesian analysis can provide the* **unconditional** *probability of being correct or incorrect about a decision from a hypothesis test (but this advantage does not come "free", it comes at the expense of having to provide your prior probability about the null hypothesis being true or not).*

(b) Since the p-value is very small, the value of $\overline{x}$ observed in the experiment must have been far from zero.

This statement is also false. A small p-value arises in this case because of a large value of the t statistic. The t statistic is typically calculated from

$$t = \frac{(\overline{x} - \mu_0)\sqrt{n}}{s}$$

Here $\mu_o = 0$, so we have

$$t = \frac{(\overline{x})\sqrt{n}}{s}$$

and hence t will be large if at least one of the following three conditions hold: $\overline{x}$ is large, or the sample size n is large, or the standard deviation s is small. Thus, while it is possible that $\overline{x}$ is far from the null value of zero, it is not necessarily the case.

2. The table below gives data on blood pressure before and after treatment for two groups of subjects participating in a clinical trial. One group took a daily calcium supplement, while the other group received a placebo.

Calcium Group		**Placebo Group**	
before	after	before	after
107	100	123	124
110	114	109	97
123	105	112	113
129	112	102	105
112	115	98	95
111	116	114	119
107	106	119	114
112	102	112	114
136	125	110	121
102	104	117	118
–	–	130	133

(a) Calculate a 95% confidence interval for the difference in blood pressure changes (before minus after) between the two groups. Give the interpretation of this confidence interval.

We will use R to calculate our confidence interval here:

```
#  First enter the data:

> calc.before <- c(107, 110, 123, 129, 112, 111, 107, 112, 136, 102)
```

```
> calc.after <- c( 100, 114, 105, 112, 115, 116, 106, 102, 125, 104)

> plac.before <- c(123, 109, 112, 102, 98, 114, 119, 112, 110, 117, 130)

> plac.after <- c(124, 97, 113, 105, 95, 119, 114, 114, 121, 118, 133)

#  Now calculate the within group differences

> calc.diff <- calc.before - calc.after

> plac.diff <- plac.before - plac.after

#  Now use the R t.test command to get
#  the required confidence interval

> t.test(calc.diff, plac.diff)

        Welch Two Sample t-test

data:  calc.diff and plac.diff t = 1.7169, df = 15.538,
p-value = 0.1059 alternative hypothesis: true difference in means is not equal
to 0 95 percent confidence interval:
 -1.339667 12.612394
sample estimates:
 mean of x  mean of y
 5.0000000 -0.6363636
```

So the confidence interval is (-1.33, 12.61). This confidence includes the null value, but also includes values that are highly clinically relevant, especially on the the side that favours calcium. Therefore, the results are inconclusive, more research is required.

(b) Carry out a t-test of the null hypothesis that there is no difference in blood pressure changes between the two groups. State the null and alternative hypotheses, calculate the test statistic, and state your conclusion.

The t-test result is already included in the above output, with $p = 0.1059$. We cannot reject the null hypothesis, but neither can we conclude that the null hypothesis is correct. As is the case with most hypothesis tests, we have not learned very much clinically useful information.

(c) You now must make a decision regarding whether or not to prescribe calcium supplementation to your patients with mild high blood pressure. In helping you to make this decision, would your answer to part (a) or (b) provide more useful information? Why?

Certainly (a), which provides a range of possible differences compatible with the data, is more useful than the p-value calculated in (b). As for decision making, it seems that in mild cases, it may be worth trying calcium for a while, and if that does not help, try another therapy. It might also be defendable to go right to another therapy more proven than calcium, as the evidence is very weak here, even though an effect from calcium cannot be ruled out.

3. The following data are observed in an experiment designed to compare a new treatment to a standard therapy:

	Therapy		
	New	Standard	
Success	60	30	90
Failure	10	40	50
	70	70	140

(a) Test the null hypothesis that there is no difference in success rates between the new and standard therapies. State the null and alternative hypotheses, and calculate a p-value using a χ^2 test. State your conclusion.

Using R:

```
> prop.test(c(60,30), c(70,70), correct=F)
```

```
        2-sample test for equality of proportions without continuity
        correction

data:  c(60, 30) out of c(70, 70)
X-squared = 28, df = 1, p-value = 1.213e-07
alternative hypothesis: two.sided
95 percent confidence interval:
 0.2865881 0.5705547
sample estimates:
   prop 1    prop 2
0.8571429 0.4285714
```

We reject the null hypothesis, $p = 0.0000001213$. The two therapies are not equally efficacious.

(b) Repeat part (a), but use a (two-sided) Fisher's Exact test instead.

Again using R:

```
> fisher.test(matrix(c(60,30,10,40), nrow=2, byrow=T))

        Fisher's Exact Test for Count Data

data:  matrix(c(60, 30, 10, 40), nrow = 2, byrow = T)
p-value = 1.67e-07
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
  3.319662 20.183569
sample estimates:
odds ratio
  7.867932
```

We reject the null hypothesis, $p = 0.000000167$. *The two therapies are not equally efficacious.*

(c) How do the p-values calculated by the two different procedures compare?

The p-values from (a) and (b) are similar but not identical.

(d) Calculate a confidence interval for the difference in success rates between the new and standard therapies. Overall, what would you conclude? After knowing the confidence interval, do the p values discussed in parts (a) or (b) add any clinically useful information? If so, what information do they add?

The confidence interval is already given in part (a), it is (0.287, 0.570). There is at least an almost 30% improvement in success rate with the new therapy, and the difference could be as large as almost 60%. Therefore, the new therapy is certainly a very substantial improvement. The p-values calculated in parts (a) and (b) add nothing of clinical utility once the confidence interval is known. Even though the p-values are very small, one cannot be sure from the p-values alone that the result is clinically meaningful (see, for example, answer to part (b) of question 1 above).

4. (a) Construct your normal prior distribution for the average age at which men get their first myocardial infarction (MI, i.e., a heart attack), among all men who do have MI's. There is no "correct" answer here, but you should justify your choice; the mean and standard deviation from this distribution should correctly represent your prior knowledge about ages at which men have heart attacks.

I know relatively little about age at first heart attacks, but reason as follows: I know that there are relatively few heart attacks before age 40, and they can occur anywhere up to the end of life, say an upper limit of around 90 years. I expect the average to not be near these extremes, but somewhere in the middle, say around 65 years. I am quite certain it should be above age 50, and below age 80, so will form a prior centered at 65, and with 95% HPD of about (50, 80). Thus, my prior would be normal(65, 58), which has 95% credible set of (50.1, 79.9). Those with better knowledge in this clinical area may have much narrower prior distributions.

(b) Update your prior to a posterior distribution using the data given below. Assume that the variance of the age at first MI is known to be 144 (so SD=12). First calculate this by hand, showing all of your calculations. Then, check your answer using the R program for Bayesian posterior means we discussed in class.

68 52 62 67 68 56 53 81 36 77

Running the program in R from the course web page, we have:

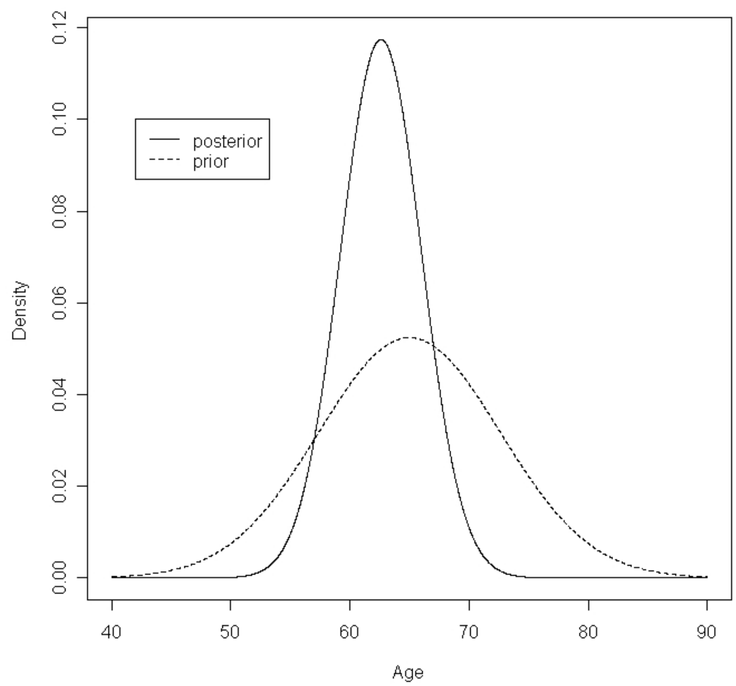
```
> x<- c(68, 52, 62, 67, 68, 56, 53, 81, 36, 77)

> post.normal.mean(x, prior.mean=65, prior.var=58, data.var=144)
$post.mean
[1] 62.59669

$post.var
[1] 11.53591
```

So the posterior distribution is N(62.60, 11.54).

(c) Use R to plot your prior and posterior distributions on the same graph. Comment on the degree to which you think your prior distribution influences the posterior distribution.

```
#  Plotting in R proceeds as follows:
x <- seq(40, 90, by=0.01)
plot(x, dnorm(x, mean=62.6, sd=sqrt(11.54)), type="l", lty=1,
     xlab="Age", ylab="Density ")
points(x, dnorm(x, mean=65, sd=sqrt(58)), type="l", lty=2)
legend(x=42, y=.1, legend=c("posterior", "prior"), lty=c(1,2))
```

The graph is:



(d) Find prior and posterior 95% credible intervals for the mean age. You can do this calculation either by hand, or using R.

Using R, we calculate:

```
#  For the posterior:

> qnorm(c(0.025, 0.975), mean=62.6, sd=sqrt(11.54))

[1] 55.94189 69.25811

#  For the prior:

> qnorm(c(0.025, 0.975), mean=65, sd=sqrt(58))
[1] 50.07336 79.92664
```

5. Suppose that a person claims that they are a psychic, and you decide to test them by having them predict the outcome of a series of coin flips that you will carry out.

(a) Suppose you flip a coin five times, and they in fact correctly predict the outcome (heads or tails) all five times. What is your personal probability that they are in fact a psychic? [Note: there is no "correct" or "incorrect" answer here, I expect all students in the class to give different answers, but there is a correct way to derive your personal probability. Therefore, provide details about where your prior probability came from.]

As the problem states, there is no single correct answer here, but the reasoning may proceed like the following:

I am personally very doubtful that psychics of this type exist (say). To quantify this, I would say that I would give odds of at least 1000 to 1 that this person is not a psychic. Thus, my prior probability is P(this person is a psychic) $\approx$ *0.001.*

Using Bayes Theorem, we now find:

$$
\begin{aligned}
& P(psychic|five\ in\ a\ row) \\
= & \frac{P(five\ in\ a\ row|psychic) * P(psychic)}{P(five\ in\ a\ row|psychic) * P(psychic) + P(five\ in\ a\ row|not\ psychic) * P(not\ psychic)} \\
= & \frac{1 \times 0.001}{1 \times 0.001 + 1/(2^5) \times 0.999} \\
= & \frac{0.001}{0.001 + 0.0312} \\
= & 0.0311
\end{aligned}
$$

Thus, I remain highly sceptical.

(b) Same question as in part (a), but re-evaluate now if they get 20 flips in a row correctly.

Using the same reasoning (and same prior) as in part (a), from Bayes Theorem we find:

$$
\begin{aligned}
& P(psychic|20\ in\ a\ row) \\
= \;& \frac{P(20\ in\ a\ row|psychic) * P(psychic)}{P(20\ in\ a\ row|psychic) * P(psychic) + P(20\ in\ a\ row|not\ psychic) * P(not\ psychic)} \\
= \;& \frac{1 \times 0.001}{1 \times 0.001 + 1/(2^{20}) \times 0.999} \\
= \;& \frac{0.001}{0.001 + 0.0000009527} \\
= \;& 0.99905
\end{aligned}
$$

At this point, despite my initial scepticism, I should be strongly re-evaluating my initial opinion.

(c) Now consider the following slightly different situation: A skeptics organization in the US offers a prize of $10 million to anyone who can prove they are a psychic, by correctly guessing the outcome of 20 coin flips in a row. Given the size of the prize, it is not surprising that the contest drew many entrants, all claiming to be psychics, and in fact, there were 1,000,000 such trials carried out on 1,000,000 different people, each claiming to be psychic. Out of these trials, one contestant actually indeed got all 20 flips correct. Do you believe their claim that they are psychic? Explain why or why not.

For any single psychic, the probability of 20 in a row (purely by chance) is $1/2^{20}) = 0.0000009537$*. However, if this trial is repeated 1,000,000 times, the probability of getting one or more "events" (i.e., 20 in a row) is given by the binomial distribution (we will calculate the binomial probability of no events, so prob of one or more is the inverse of this),*

$$1 - \frac{1000000!}{0!1000000!}(0.0000009537)^0 \times (1 - 0.0000009537)^{1000000} = 1 - 0.3853224 = 0.6146776$$

Now returning to Bayes Theorem once again, we find

$$
\begin{aligned}
& P(psychic|20\ in\ a\ row\ in\ million\ trials) \\
= \;& \frac{P(20\ in\ a\ row\ in\ million\ trials|psychic) * P(psychic)}{P(20\ in\ a\ row\ in\ million\ trials|psychic) * P(psychic) + P(20\ in\ a\ row\ in\ million\ trials|not\ psychic) * P(not\ psychic)} \\
= \;& \frac{1 \times 0.001}{1 \times 0.001 + 0.6147 \times 0.999}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{0.001}{0.001 + 0.6140853} \\
&= 0.0016258
\end{aligned}
$$

Thus, my prior probability is hardly changed if one or more psychics get 20 in a row out of one million trials of this experiment.

Conclusion: Events that are rare still do happen, every once in a while. If you keep trying over and over again, sooner or later a rare event will arise by chance. If you test a single "psychic" once, 20 in a row would be pretty convincing evidence. Twenty in a row for one out of 1,000,000 "psychics", however, is a very common event, in fact happening more than half the time.