Exercise 3.1

i) This obstetrician has performed 20 deliveries (n = 20); each delivery has two possible outcomes, either delivery by Caesarean Section (CS) or not. Let R be the random variable that denotes the number of patients delivered by CS for this obstetrician, $R \sim \text{Bin}(n, \pi)$. Let r be the value taken by R in the sample. Here r = 8. He is interested in testing whether his probability of delivering a patient by CS, π, is equal to that in the population, *i.e.* he would like to test the null hypothesis: $H_0: \pi = 0.20$ against $H_A: \pi \neq 0.20$.

ii) What values of R would be as or more extreme than the observed r=8. Under the null hypothesis $E[R] = n\pi = 20 \times 0.2 = 4$. The observed value r = 8. Possible values of R that are as or more extreme under the null is given by $|R - 4| \geq |8 - 4|$. Values of R for which this is satisfied are {0, 8, 9, … , 20}.

From Neave's binomial tables where n=20 and p=0.2 we can see that P(R=8)=0.0222. Looking at the other possible outcomes we see that R=0, 9, 10, … , 20 are all more extreme in probability terms .

The binomial distribution: individual probabilities

Prob ($X = x$)

n	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10	.15	⋮	.20	.25	.30	⋮	.35	.40	.45	.50
$n = 20$																				
0	.8179	.6676	.5438	.4420	.3585	.2901	.2342	.1887	.1516	.1216	.0388	.0261	.0115	.0032	.0008	.0003	.0002	.0000	.0000	.0000
1	.1652	.2725	.3364	.3683	.3774	.3703	.3526	.3282	.3000	.2702	.1368	.1043	.0576	.0211	.0068	.0030	.0020	.0005	.0001	.0000
2	.0159	.0528	.0988	.1458	.1887	.2246	.2521	.2711	.2818	.2852	.2293	.1982	.1369	.0669	.0278	.0143	.0100	.0031	.0008	.0002
3	.0010	.0065	.0183	.0364	.0596	.0860	.1139	.1414	.1672	.1901	.2428	.2379	.2054	.1339	.0716	.0429	.0323	.0123	.0040	.0011
4	.0000	.0006	.0024	.0065	.0133	.0233	.0364	.0523	.0703	.0898	.1821	.2022	.2182	.1897	.1304	.0911	.0738	.0350	.0139	.0046
5	.0000	.0000	.0002	.0009	.0022	.0048	.0088	.0145	.0222	.0319	.1028	.1294	.1746	.2023	.1789	.1457	.1272	.0746	.0365	.0148
6	.0000	.0000	.0000	.0001	.0003	.0008	.0017	.0032	.0055	.0089	.0454	.0647	.1091	.1686	.1916	.1821	.1712	.1244	.0746	.0370
7	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0005	.0011	.0020	.0160	.0259	.0545	.1124	.1643	.1821	.1844	.1659	.1221	.0739
8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0004	.0046	.0084	.0222	.0609	.1144	.1480	.1614	.1797	.1623	.1201
9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0011	.0022	.0074	.0271	.0654	.0987	.1158	.1597	.1771	.1602
10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0005	.0020	.0099	.0308	.0543	.0686	.1171	.1593	.1762
11	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0030	.0120	.0247	.0336	.0710	.1185	.1602
12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0008	.0039	.0092	.0136	.0355	.0727	.1201
13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0010	.0028	.0045	.0146	.0366	.0739
14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0007	.0012	.0049	.0150	.0370
15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0013	.0049	.0148
16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0046
17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011
18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002
19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
20	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	.99	.98	.97	.96	.95	.94	.93	.92	.91	.90	.85	⋮	.80	.75	.70	⋮	.65	.60	.55	.50

Prob ($X = x$)

iii) *Calculate p-value using n* = 20, *r* = 8 and *p* = 0.2

Recall that the p-value is the probability of observing a sample as or more extreme than that observed given that the null hypothesis is true.

Since $r = 8 > E[R] = n\pi = 20 \times 0.2 = 4$, the **one-sided** p-value = $P_8 + P_9 + \ldots + P_{20}$.

From the table above this is equal to 0.0222+0.0074+0.0020+0.0005+0.0000+….=0.0321 (or could use the cumulative binomial table). Adopting the approach of doubling the 1-sided value gives a 2-sided p-value = 0.0642.

Alternatively we could calculate the two sided *p*-value by adding the probabilities that are at least as small as P_8 in the opposite tail of the sampling distribution. We saw in (ii) above that this includes the value R=0. In the binomial table we see the P(0)=0.0115. Hence the **two-sided** *p*-value = $P_0 + P_8 + P_9 + \ldots + P_{20}$ = 0.0115 + 0.0321 = 0.044

The fact that the *p*-values from the two approaches fall either side of 0.05 illustrates the importance of deciding which statistical approach to adopt **before** carrying out the analysis.

If it had been decided to simply double the 1-sided *p*-value the result should be described as a borderline statistically significant result. Formally, since $p>0.05$, we do not have evidence that the obstetrician's rate differs from the national average rate of 20%, but the borderline nature of the result would be worth noting in a description of the result. As is always the case with non-statistically significant results ***it would not be correct to say that this provides evidence that the obstetrician's rate is equal to the national average,*** merely (the weaker statement) that it does not provide evidence against this hypothesis.

If it had been decided to add the probabilities in the opposite tail of the sampling distribution that are at least as small as that for the observed result it would still be advisable to describe the result as being of borderline statistical significance. Formally, since $p<0.05$, we do have evidence that the obstetrician's rate differs from the national average rate of 20%, but the borderline nature of the result would still be worth noting.

This version of the test can also be performed using Stata.

```
. bitesti 20 8 0.2

 N  Observed k   Expected k   Assumed p   Observed p
------------------------------------------------------------
20          8            4       0.20000      0.40000

Pr(k >= 8)            = 0.032143  (one-sided test)
Pr(k <= 8)            = 0.990018  (one-sided test)
Pr(k <= 0 or k >= 8)  = 0.043672  (two-sided test)
```

iv) There are many problems with this experiment. These include:

a) A London teaching hospital might have different types of patients from those involved in the calculation of the national average rate.
b) The CS rate at the weekend might be expected to be different from that during the week.
c) The small sample size in this study means that the statistical power to detect moderate differences is low. This is illustrated by the fact that although the obstetrician's rate is twice the national average, this difference is not statistically significant.
d) The national average includes the experience of both junior and senior doctors *etc.*

Exercise 3.2

i) In this study in the Burnley area n = 286, R takes the value r =168 and the estimator of the proportion of smokers (P) takes the value p = $\frac{168}{286} = 0.587$. We wish to assess the evidence that the proportion of smokers in the Burnley area differs from the national average. Formally we wish to test H_0: $\pi = 0.3$ against H_1: $\pi \neq 0.3$.

With a large sample size, we can use the normal approximation to the binomial distribution. Given that H_0 is true, the following random value approximately follows a N(0,1) distribution:

$$Z = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{P - 0.30}{\sqrt{\frac{0.30 \times 0.70}{286}}}.$$

Since p = 0.587 > $\pi_0 = 0.30$, the **one-sided** p-value is given by:

Prob (Z > $\frac{0.587 - 0.3}{\sqrt{\frac{0.3 \times 0.7}{286}}} = 10.6$) < 0.0001

The p-value is so small that even when doubling this we still get a two-sided p<0.0001. The p-value alone only tells us that there is strong evidence against the null hypothesis i.e. evidence that the proportion of smokers in the Burnley area is different than the national average. Taking into consideration both the sample proportion 0.587 which is greater than 0.3 and that p<0.0001 we can conclude there is strong evidence that the proportion of smokers in the Burnley area is greater than the national average.

ii) To construct a 95% CI we use the equation from Case Study 3 in Analytical Techniques 2. The estimator of the 95% CI is: $\mathrm{P} \pm 1.96\sqrt{\frac{\mathrm{P}(1-\mathrm{P})}{\mathrm{n}}}$

$$\text{which takes the value} = 0.587 \pm 1.96\sqrt{\frac{0.587(1-0.587)}{286}}$$
$$= (0.530, 0.644).$$

iii) Due to the link between hypothesis tests and confidence intervals we can infer from the fact that the 95% confidence interval in ii) does not include 0.3 that the result of the hypothesis test in i) which tests the null hypothesis that π = 0.3 is statistically significant at the 5% level.

Occasional violations of this rule can occur in practice. Indeed this is theoretically possible for the formulae here since the variance of $p - \pi_0$ is (correctly) estimated using different formulae when constructing confidence intervals ($p(1-p)/n$) and carrying out hypotheses tests ($\pi_0(1-\pi_0)/n$). In practice confusion rarely arises from this.]

Exercise 3.3 Stata output and commentary

```
. use vit_e , clear

. sort set case

. list if set<=5, sep(2)
```

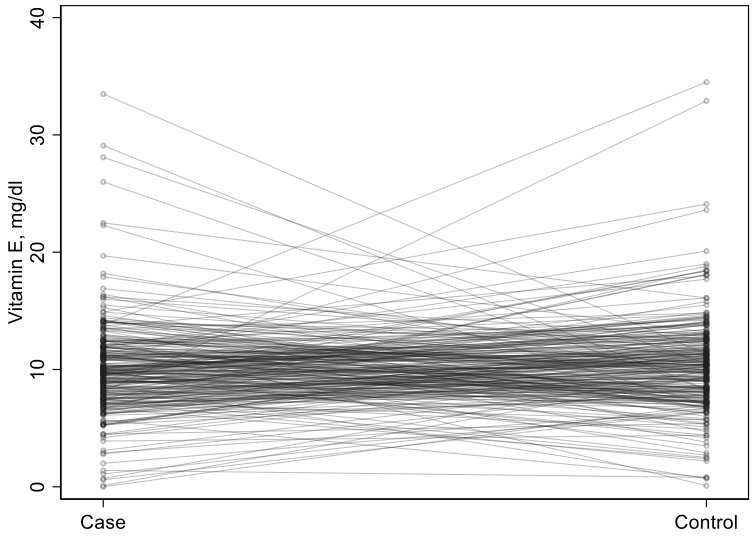
	set	case	vit_e
1.	1	Case	.6
2.	1	Control	6.4
3.	2	Case	5.3
4.	2	Control	12.1
5.	3	Case	13.3
6.	3	Control	3.8
7.	4	Case	12.5
8.	4	Control	12.9
9.	5	Case	15.3
10.	5	Control	4.7

} data from first matched pair

} data from second matched pair

Note: data from cases and controls is on separate lines

```
twoway connected vit_e case  ,
      c(L) sort(set case)
      xlab(1 2 , value)
      xscale(range(0.95 2.05))  xtitle("")
      lwidth(0.5pt) lcol(gs2%30)
      msym(oh) msize(3pt) mcol(gs2%30)
      scheme(s1mono)
```



The large amount of data makes it difficult to see any pattern here. This type of graph has limited utility with 'large' datasets.

```
. reshape wide vit_e , i(set) j(case)
(j = 1 2)

Data                               Long   ->   Wide
-----------------------------------------------------------------------------
Number of observations              542   ->   271
Number of variables                   3   ->   3
j variable (2 values)              case   ->   (dropped)
xij variables:
                                  vit_e   ->   vit_e1 vit_e2
-----------------------------------------------------------------------------
```

```
. list in 1/5
```

	set	vit_e1	vit_e2
1.	1	.6	6.4
2.	2	5.3	12.1
3.	3	13.3	3.8
4.	4	12.5	12.9
5.	5	15.3	4.7

```
label var vit_e1 "Vitamin E Cases"
label var vit_e2 "Vitamin E Controls"
```

iv) paired t-test

```
. ttest vit_e1=vit_e2

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
  vit_e1 |     271    10.03063    .2473182    4.071371     9.54371    10.51754
  vit_e2 |     271    10.31292    .2483331    4.088079       9.824    10.80183
---------+--------------------------------------------------------------------
    diff |     271   -.2822878    .3306562    5.443288   -.9332802    .3687045
------------------------------------------------------------------------------
     mean(diff) = mean(vit_e1 - vit_e2)                            t =  -0.8537
 H0: mean(diff) = 0                           Degrees of freedom =      270

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.1970         Pr(|T| > |t|) = 0.3940          Pr(T > t) = 0.8030
```

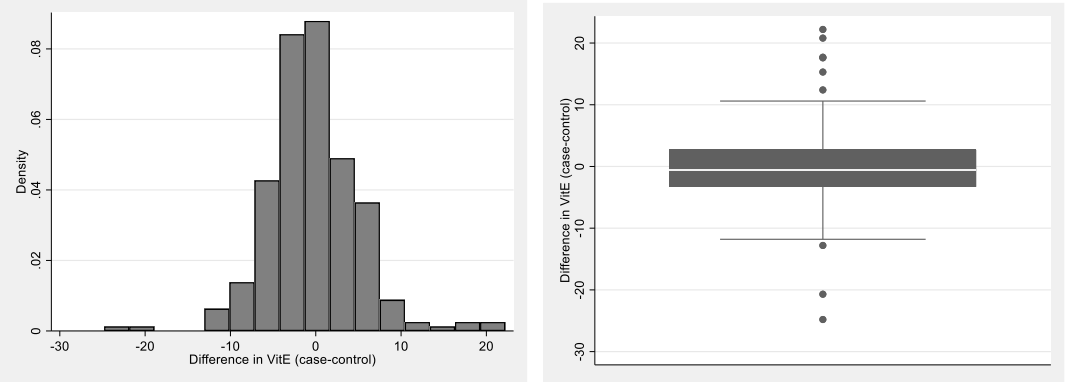
The estimated mean difference, 95% confidence interval, t-statistic and two-sided p-value are highlighted in bold font.

Interpretation: 'On average cases had lower vitamin E levels than controls, but this difference was not statistically significant (p-value = 0.39). The estimated mean difference was -0.28mg/dl (95% CI -0.93, 0.37 mg/dl).'

Note: when describing results try and mention i) the direction, ii) the magnitude and iii) the precision of estimates as well as (or instead of) their statistical significance.

(v) The test assumes that the **differences** in Vitamin E levels between a cases and its control are i) independent and ii) follow a normal distribution.

```
. gen diff= vit_e1 – vit_e2
. label var  diff "vit E: case control difference"
. histogram diff
. graph box diff,
```



Two ways of displaying the distribution of case-control differences

The distribution of the case-control differences does not differ markedly from that of a normal distribution. This, coupled with the relatively large sample size, means that we can be satisfied with the reliability of our inferences in part iv).