18.05 Problem Set 8, Spring 2014 Solutions

Problem 1. (10 pts.) **(a)** Let $x = $ number of heads
Model: $x \sim \text{binomial}(12, \theta)$.
Null distribution binomial$(12, 0.5)$.
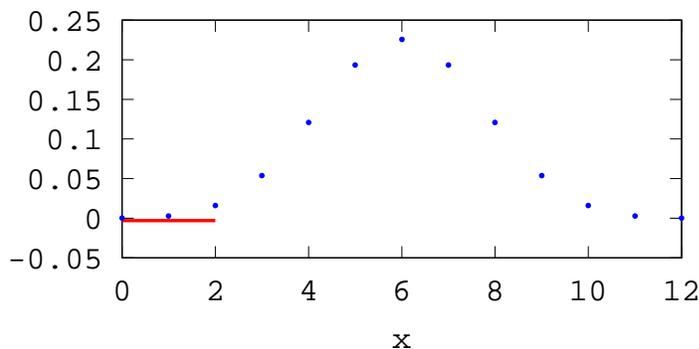Data: 3 heads in 12 tosses.

Since H_A is one-sided the rejection region is one-sided. Since H_A says that θ is small it predicts a small number of heads in 12 tosses. That is, we reject H_0 on a small number of heads.

So, rejection region = left tail of null distribution.

$$c_{0.95} = \texttt{qbinom(0.05, 12, 0.5)} - 1 = 2$$

Rejection region is $0 \le x \le 2$.

$$p = \texttt{pbinom(3, 12, 0.5)} = 0.072998$$

Binomial(12,15) null distribution and rejection region $x \le 2$.

Erika concludes there is not enough evidence to reject the null hypothesis at the significance level 0.05.

(b) Let $n = $ number of tosses that were tails before the third that is heads
Probability model: Choose two tosses in the first n+2 for heads; the $n + 3$rd toss must be heads:

$$p(n) = \binom{n+2}{2}(1-\theta)^n\theta^3.$$

This is called the negative binomial distribution with parameters 3 and θ. Data: 9 tails to get 3 heads.

Since H_A is one-sided we the rejection region is one-sided. Since H_A is that θ is small it predicts a large number of tails before 3 heads. So we reject on a large number of tails.

Rejection region = right tail of null distribution.

$$c_{0.05} = \texttt{qnbinom(0.95, 3, 0.5)} + 1 = 9$$

Rejection region is $n \geq 9$.

$$p = \texttt{1 - pnbinom(8, 3, 0.5)} = 0.032715$$
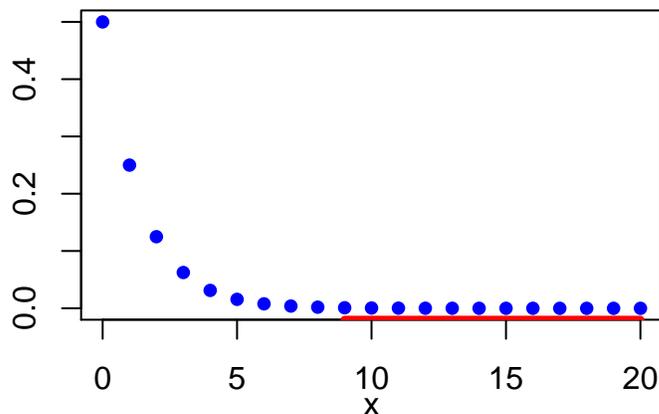
Negative binomial null distribution and rejection region

Ruthi rejects the null hypothesis in favor of H_A at significance level 0.05.

(c) No. Computing a p-value requires that the experiment be fully specified ahead of time so that the definition of 'data at least as extreme' is clear.

(d) Prior: beta(n, m) has pdf $c\, \theta^{n-1}(1 - \theta)^{m-1}$

Likelihood experiment 1: $\binom{12}{3} \theta^3 (1 - \theta)^9$

Likelihood experiment 2: $\binom{11}{2} \theta^3 (1 - \theta)^9$

Since the likelihoods are the same up to a constant factor the posterior has the same form

$$c\, \theta^{n+3-1}(1 - \theta)^{m+9-1}$$

which is the pdf of a beta$(n + 3, m + 9)$ distribution.

The two posteriors are identical. In the Bayesian framework the same data produces the same posterior.

(e) The main point is that in the frequentist framework the decision to reject or accept H_0 depends on the exact experimental design because it uses the probabilities of unseen data as well as those of the actually observed data.

Problem 2. (10 pts.) We use the χ^2 test for $\sigma^2 = 1$.

Call the data x.
$\sigma^2 = 1$
$s^2 = $ sample variance $= \texttt{var(data)} = 2.3397$
$n = 12 = $ number of data points
χ^2-statistic: $X^2 = (n - 1)s^2/\sigma^2 = 25.7372$
The one-sided p-value is $\texttt{p = 1 - pchisq(}X^2\texttt{, n-1) = 0.0071}$.

Since $p < 0.05$ we reject the null hypothesis H_0 that $\sigma^2 = 1$ in favor of the alternative that $\sigma^2 > 1$.

Problem 3. (10 pts.)   We do a χ^2 test of goodness of fit comparing the observed counts with the counts expected from Benford's distribution.

You can use either test statistic

$$G = 2 \sum O_i \ln \left(\frac{O_i}{E_i} \right).$$

or

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where O_i are the observed counts and E_i are the expected counts from Benford's distribution. The total count $= 100$.

First digit k	1	2	3	4	5	6	7	8	9
observed	7	13	12	9	9	13	11	10	16
expected	30.103	17.609	12.494	9.691	7.918	6.695	5.7992	5.1153	4.5757
X^2 components	17.731	1.206	0.200	0.049	0.148	5.939	4.664	4.665	28.523

The χ^2-statistics are $G = 56.3919$ and $X^2 = 62.6998$.
There are 9 cells that must sum to 100 so the degrees of freedom $= 8$.
The p-value using G is

$$p = P(\text{G test stat} > 56.3919 \,|\, H_0) = \texttt{1-pchisq(56.3919, 8)} = 2.4 \times 10^{-9}$$

The p-value using X^2

$$p = P(\text{X}^2 \text{ test stat} > 62.6998 \,|\, H_0) = \texttt{1-pchisq(62.6998, 8)} = 1.4 \times 10^{-10}$$

Since $p < \alpha$ we reject H_0 in favor of the notion that Jon and Jerry were trying to embezzle money.

Problem 4. (10 pts.)

(a) We let $x =$ the first set of 20 numbers and y the second. R makes it almost too easy. We give the command `var.test(x,y)` R then prints out

```
data:  x and y
F = 0.9703, num df = 19, denom df = 19, p-value = 0.9484
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3840737 2.4515249
sample estimates:
ratio of variances
        0.9703434
```

The p-value is 0.9484 with F-statistic 0.9703.

(b) We found the formula for the F statistic for this test at
http://en.wikipedia.org/wiki/F-test_of_equality_of_variances

$$s_x^2 = \texttt{var(x)} = 1.1302$$
$$s_y^2 = \texttt{var(y)} = 1.1647$$

Our F-statistic is

$$\texttt{fstat} = \frac{s_x^2}{s_y^2} = 0.9703$$

The degrees are freedom are both 19. Since the F-statistic is less than 1, the p-value is

$$p = \texttt{2*pf(fstat, 19, 19))} = 0.9484$$

which matches our result in part (a).

Problem 5. (10 pts.)   **(a)**   Let's specify the assumptions and hypotheses for this test.
We have 4 groups of data: Clean, 5-day, 10-day, full
Assumptions: Each group of data is drawn from a normal distribution with the same variance σ^2; all data is drawn independently.
H_0: the means of all the normal distributions are equal.
H_A: not all the means are equal.

The test compares the between group variance with the within group variance. Under the null hypothesis both are estimates of σ^2, so their ratio should be about 1. We'll reject H_0 if this ratio is far from 1.

We used R to do the computation. Here's the code.

```
mns = c(1.32, 1.26, 1.53, 1.39)
v = c(0.56, 0.80, 0.93, 0.82)
m = 351 # number of samples per group
n = length(mns) # number of groups
msb = m*var(mns) # between group variance
msw = mean(v) # within group variance
fstat = msb/msw
df1 = n-1;
df2 = n*(m-1)
p = 1 - pf(fstat, df1,df2)

print(fstat)
print(p)
```

This produced an F-statistic of 6.09453 and $p = 0.00041$. Since the p-value is much smaller than 0.05 we reject H_0.

(b) To compare all 4 means 2 at time would require $\binom{4}{2} = 6$ t-tests. If we run six tests it is not appropriate to claim the significance level of each one is the significance level of the collection.

(c) We compare 10-day beards with each of the others. In each case we have: H_0: the means are the same
H_A: the 10-day mean is greater than the other mean.

Note carefully that this is a one-sided test while the F-test in part (b) is a two-sided test.

From the class 19 reading we have the t-statistic for two samples. Since both samples have the same size $m = 351$ the formula looks a little simpler.

$$t = \frac{\bar{x} - \bar{y}}{s_{\bar{x}-\bar{y}}},$$

where the pooled sample variance is

$$s_P^2 = \frac{s_x^2 + s_y^2}{m}$$

Note: the test assumes equal variances which we should verify in each case. This raises the issue of multiple tests from the same data, but it is legitimate to do this as exploratory analyis which merely suggests directions for further study.

The following table shows the one-sided, 2-sample t-test comparing the mean of the 10-day growth against the other three states.

	t-stat	one-sided p-value	F-stat
clean	3.22314	0.00066	10.38866
5-day	3.84587	0.00007	14.79069
full	1.98273	0.02389	3.93120

We also give the F-statistic for the two samples. You can check that the F-statistic for two-samples is just the square of the t-statistic.

We reiterate, with multiple testing the true significance level of the test is larger than the significance level for each individual test.

18.05 Introduction to Probability and Statistics

Spring 2014