The University of Sydney
STAT2012 Statistical Tests

Semester 2	Computer Practice Week 6	2015

Useful R commands

- The R codes to calculate the values related to χ^2_n, i.e. $\Pr(\chi^2_n \le q) = p$ and $F_{m,n}$, i.e. $\Pr(F_{m,n} \le q) = p$ are

```
dchisq(x,n),  pchisq(q,n),  qchisq(p,n),  rchisq(r,n);
df(x,m,n),    pf(q,m,n),    qf(p,m,n),    rf(r,m,n).
```

- *Chi-square test:* To test $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 :\ \sigma^2 \neq \sigma_0^2$ (two-sided) or $H_1 :\ \sigma^2 > \sigma_0^2$ or $H_1 :\ \sigma^2 < \sigma_0^2$ (one-sided), the R codes are

```
chi20=(n-1)*s2/sigma20
chi20
p.value=1-pchisq(chi20,n-1) #upper-sided or p.value=pchisq(chi20,n-1,lower.tail=F)
p.value=pchisq(chi20,n-1)     #lower-sided
p.value=2*min(1-pchisq(chi20,n-1),pchisq(chi20,n-1))  #2-sided
p.value
```

- *F-test:* To test $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$ (two-sided) or $H_1 : \sigma_1^2 > \sigma_2^2$ or $H_1 : \sigma_1^2 < \sigma_2^2$ (one-sided), the R codes are

```
var.test(x,y,alternative="greater")  #var(x) > var(y) or swap x & y
f0=var(x)/var(y) #checking
f0
m=length(x)
n=length(y)
p.value=1-pf(f0,m-1,n-1)    #upper-sided or p.value=pf(f0,m-1,n-1,lower.tail=F)
p.value=2*(1-pf(f0, m-1,n-1))  #2-sided
p.value
```

- ANOVA test: To test $H_0 :\ \mu_1 = \mu_2 = \cdots = \mu_g$ against H_1: at least one equality does not hold, the R codes are

```
f=factor(rep(letters[1:g],c(n1,n2,n3,ng)))
f
aov.x=aov(x~f)
summary(aov.x)
```

Important points

- You will perform one sample Chi-square test on variance, two samples F-test on variances and the ANOVA test to compare the means of $g > 2$ populations.
- Only some of the R codes are provided.

Practice Problems

Open the data set `survey` containing measurements of the following variables from 95 students:

sex	1=male; 2=female
age	Year
height:	Inches
credit:	Number of credit cards in possession
pulse:	Number of heartbeats in one minute
pulse.ex:	Number of heartbeats in one minute after regular exercise over a period
exercise:	Number of hours during last week
smoke:	1=yes; 2=no
hand:	1=left-handed; 2=right-handed; 3=ambidextrous

Read the data `survey`. Set `pulse.sf` to contain the `pulse` among female students who smoke, `pulse.nf` to contain the `pulse` among female students who do not smoke, `pulse.sm` to contain the `pulse` among male students who smoke and `pulse.nm` to contain the `pulse` among male students who do not smoke.

```
survey=read.csv("http://www.maths.usyd.edu.au/u/UG/IM/STAT2012/r/survey.csv")
attach(survey)
psf=pulse[smoke==1 & sex==2]
psf
pnf=pulse[smoke==2 & sex==2]
pnf
psm=pulse[smoke==1 & sex==1]
psm
pnm=pulse[smoke==2 & sex==1]
pnm
```

Note that you can use this link to read the data set at home.

1. Test if the variance of the pulses among female students who smoke is less than 100 using the *Chi-square test*. Note that the normality assumption for the distribution of pulse rates among female smokers was checked in week 3.
   (a) State the null and alternative hypotheses.
   (b) Perform the test assuming normality for the population of female smokers. Draw your conclusion about H_0 based on the p-value.

2. Test the equality of variance assumption in the two sample t-test last week, that is, test if the variances of pulses among female smokers and non-smokers are equal using the F *test*. Note that the normality assumptions for the distributions of pulse rates among female smokers and among female nonsmokers were checked last week.

(a) State the null and alternative hypotheses.

(b) Perform the test assuming normality for the population of female smokers and non-smokers. Can the equality of variances be assumed in the two sample t-test last week?

(c) Check the p-value of the test using some R codes.

3. Test if the means of the pulses among the four groups of students: MS (male smoker), FS (female smoker), MN (male non-smoker) and FN (female non-smoker) are equal using the *ANOVA test*.

   (a) State the null and alternative hypotheses.

   (b) Perform the test and report the test statistic and p-value. Draw your conclusion about H_0 based on the p-value.

   (c) Draw four boxplots side by side and the normal qq plot of the data and comment the normality and equality-of-variance assumptions.

   (d) Check the p-value of the test using some R codes.