Stat 312: Lecture 22
Inference on Correlation

Moo K. Chung
mchung@stat.wisc.edu

April 17, 2003

Concepts

1. Assume that X and Y are two random variables whose joint probability density function f is the *bivariate normal* such that $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ and $\rho(X,Y) = \rho$. In this situation, we can show that $\mathbb{E}(Y|X = x) = \beta_0 + \beta_1 x$, where $\beta_1 = \rho\sigma_Y/\sigma_X$. This is the regression model we studied!

2. Testing $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$ is equivalent to $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$. So the appropriate test statistic for testing $H_0 : \rho = 0$ is $T = \hat{\beta}_1/S_{\hat{\beta}_1} \sim t_{n-2}$ (see Lecture 19, Concept 3). This test statistic is equivalent to $T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$.

3. Testing for independence is equivalent to testing $\rho = 0$ in bivariate data.

In-class problems

Exercise 12.67. A sample of 10,000 (x_i, y_i) pairs resulted in the sample correlation coefficient $r = 0.022$. Test $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$ at level 0.05. Also est if x_i and y_i are independent.
Solution.

```
> r<-0.022
> r*sqrt(10000-2)/sqrt(1-r^2)
[1] 2.200313
> pt(2.2,9998)
[1] 0.9860852
```

Using Concept 2, $t = 2.2$. P-value $= P(|T| > 2.2) = 2(1 - 0.986) = 0.028$. So we reject H_0. Since independence is equivalent to zero correlation in bivariate data, we can use the same test procedure and end up with the same result.

Example. Test for the dependence of the mathematical achievement test scores (x_i) and calculus grades (y_i) for 10 college freshmen assuming (x_i, y_i) are from bivariate normal.
x$\|39, 43, 21, 64, 57, 47, 28, 75, 34, 52$
y$\|65, 78, 52, 82, 92, 89, 73, 98, 56, 75$

Solution. There are two ways to solve this problem. The first approach might be much simpler conceptually. The only thing you need is to compute the correlation coefficient and follow the solution of Exercise 12.67.

```
> r<-cor(x,y)
[1] 0.8397859
> r*sqrt(10-2)/sqrt(1-r^2)
[1] 4.375015
```

An alternate method would be to do a simple linear regression analysis.

```
> summary(lm(y~x))
Coefficients:
Estimate  t value Pr(>|t|)
   x          4.4  0.0024
```

Self-study problems

Please read P.533. Example 12.16.

Homework 7

Due April 29. 11:00am.
12.30., 12.36., 12.58., 12.60.,12.66.