Week 5 Lab

Tuesday, February 13/Thursday, February 15

Plan for today

1. Two-sample proportion tests
2. Pearson's chi-square test
3. Fisher's exact test
4. McNemar's test
5. Kolmogorov-Smirnov Test
6. qqplots

Recap

Recap

- Risk ratio
- Odds ratio
- Covariances
- Correlation

Two-sample proportion tests

Overview of two-sample proportion tests

- Scenario: Paired binary data, e.g. the disease and smoke conditions of a study participant.
- Mathematical notation:
  - $(X_1, \ldots, X_n)$ is a sequence of iid random variables of $Bern(p_X)$. Denote as X^n.
  - $(Y_1, \ldots, Y_n)$ is a sequence of iid random variables of $Bern(p_Y)$. Denote as Y^n.
  - $((X_1, Y_1), \ldots, (X_n, Y_n))$ is a sequence of paired X^n and Y^n. Denote as $(X, Y)^n$.
- Data representation: 2 x 2 contingency table

2 x 2 contingency table

	$Y = 1$	$Y = 0$	Marginal Total of X
$X = 1$	$o_{1,1}$	$o_{1,2}$	$o_{1,1} + o_{1,2} = n_{1,*}$
$X = 0$	$o_{2,1}$	$o_{2,2}$	$o_{2,1} + o_{2,2} = n_{2,*}$
Marginal Total of Y	$o_{1,1} + o_{2,1} = n_{*,1}$	$o_{1,2} + o_{2,2} = n_{*,2}$	

The four entries $o_{1,1}$, $o_{1,2}$, $o_{2,1}$, and $o_{2,2}$ are the observed counts of $(X, Y)^n$. The sum of them is the sample size n.

The column and row sums are called marginal totals.

Null hypotheses

- There are mainly three ways to put the same H_0:
  - X and Y are independent. H_0: $X \perp\!\!\!\perp Y$.
  - The odds ratio is equal to 1. H_0: $(o_{1,1}/o_{1,2})/(o_{2,1}/o_{2,2}) = 1$.
  - The proportions of two samples are equal. H_0: $o_{1,1}/n_{1,*} = o_{2,1}/n_{2,*}$.
- H_0: $X \perp\!\!\!\perp Y$ is directly used to derive the test statistic and calculate p-value, but it is equalvalent to the other two H_0s.

Pearson’s chi-square test

Pearson’s chi-square test

```
chisq.test(ds_2x2_tbl, correct = F)
##
##  Pearson's Chi-squared test
##
## data:  ds_2x2_tbl
## X-squared = 14.613, df = 1, p-value = 0.000132
```

Fisher’s exact test

Fisher’s exact test

```
fisher.test(ds_2x2_tbl, alternative='two.sided')
##
##  Fisher's Exact Test for Count Data
##
## data:  ds_2x2_tbl
## p-value = 0.0001724
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.2584247 0.6776792
## sample estimates:
## odds ratio
##   0.420428
```

McNemar’s test

McNemar's test

```
data <- matrix(c(6, 2, 8, 4), ncol=2, byrow=T)
mcnemar.test(data)
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  data
## McNemar's chi-squared = 2.5, df = 1, p-value = 0.1138
```

Kolmogorov-Smirnov Test

Kolmogorov-Smirnov Test

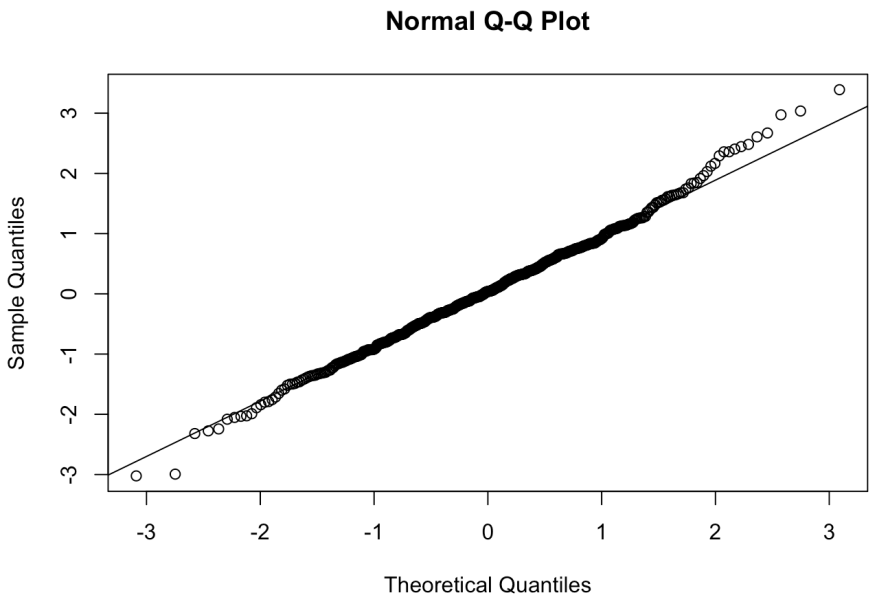
```
x <- rnorm(50)
y <- runif(30)

ks.test(x,y)
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  x and y
## D = 0.56, p-value = 6.303e-06
## alternative hypothesis: two-sided
```

qqplots

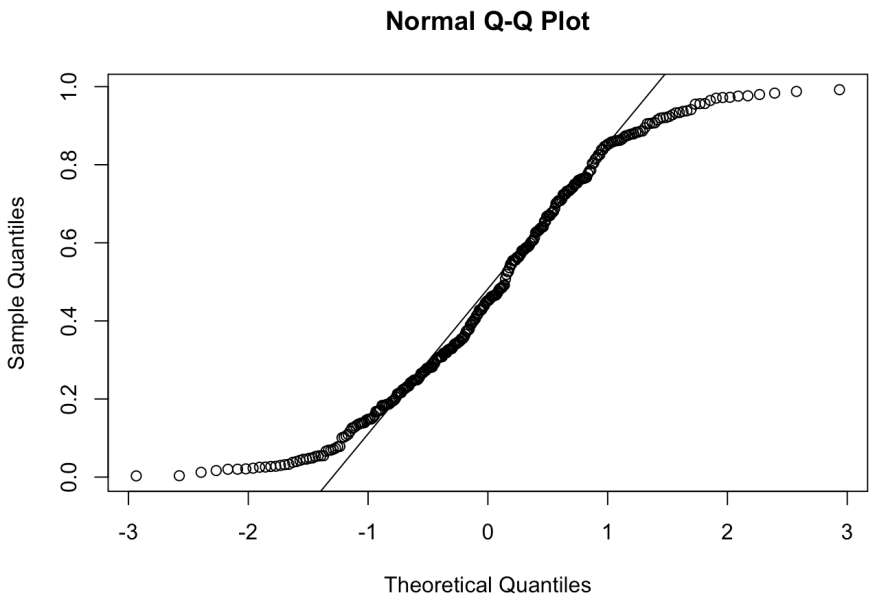
qqplots

```
x <- rnorm(500)
qqnorm(x)
qqline(x)
```



qqplots

```
y <- runif(300)
qqnorm(y)
qqline(y)
```



Questions?