GV300 - Quantitative Political Analysis

University of Essex - Department of Government

Lorenzo Crippa

Week 18 – 27 January, 2020

What are we doing today?

Today's session:

1. Robust estimation and Generalized Least Squares (GLS)
2. Non-parametric tests

Robust estimation and Generalized Least Squares (GLS)

The importance of constant variance for OLS

Remember vcov matrix, OLS estimates of parameters and S.E.:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N - K - 1}$$

$$\widehat{Var}(\hat{\beta}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

The importance of constant variance for OLS

Remember vcov matrix, OLS estimates of parameters and S.E.:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N - K - 1}$$

$$\widehat{Var}(\hat{\beta}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

Assumption 5: Homoscedasticity and no serial correlation

1. $Var(u_i|\mathbf{X}) = \sigma^2$ for $i = 1, 2, \ldots, N$

The importance of constant variance for OLS

Remember vcov matrix, OLS estimates of parameters and S.E.:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N - K - 1}$$

$$\widehat{Var}(\hat{\beta}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

Assumption 5: Homoscedasticity and no serial correlation

1. $Var(u_i|\mathbf{X}) = \sigma^2$ for $i = 1, 2, \ldots, N$
2. $Cov(u_i, u_j|\mathbf{X}) = 0$ for all $i \neq j$

The importance of constant variance for OLS

Remember vcov matrix, OLS estimates of parameters and S.E.:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N - K - 1}$$

$$\widehat{Var}(\hat{\beta}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

Assumption 5: Homoscedasticity and no serial correlation

1. $Var(u_i|\mathbf{X}) = \sigma^2$ for $i = 1, 2, \ldots, N$
2. $Cov(u_i, u_j|\mathbf{X}) = 0$ for all $i \neq j$

What if we have heteroscedasticity and/or serial correlation or autocorrelation?

Some matrix algebra

Let's work on the vcov matrix:

$$
\begin{aligned}
&\widehat{Var}(\hat{\beta}) \\
&\qquad = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \\
&\qquad = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&\qquad = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\sigma}^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&\qquad = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\hat{\mathbf{u}}\hat{\mathbf{u}}'|\mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}
\tag{1}
$$

Some matrix algebra

Let's work on the vcov matrix:

$$
\begin{aligned}
&\widehat{Var}(\hat{\beta}) \\
&\quad = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \\
&\quad = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&\quad = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\sigma}^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&\quad = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\hat{\mathbf{u}}\hat{\mathbf{u}}'|\mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned} \tag{1}
$$

With:

$$
E(\hat{\mathbf{u}}\hat{\mathbf{u}}'|\mathbf{X}) = \begin{bmatrix} \hat{\sigma}^2 & 0 & \dots & 0 \\ 0 & \hat{\sigma}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\sigma}^2 \end{bmatrix} = \hat{\sigma}^2\mathbf{I} = \Sigma
$$

Some matrix algebra

Let's work on the vcov matrix:

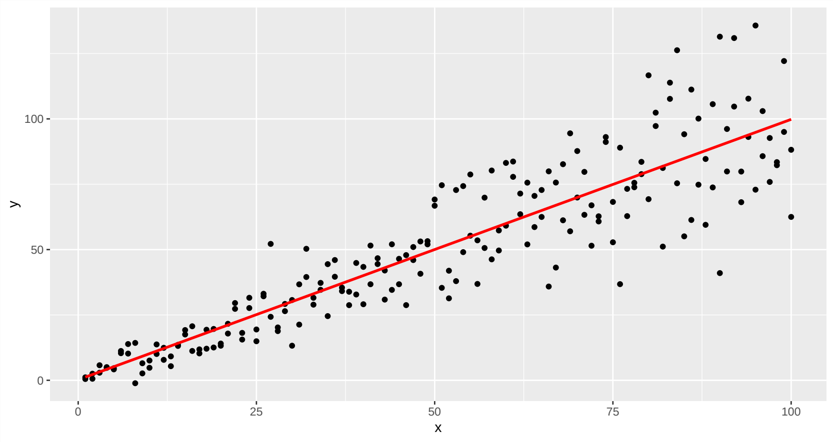
$$
\begin{aligned}
\widehat{Var}(\hat{\beta}) & \\
&= \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \\
&= \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\sigma}^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\hat{\mathbf{u}}\hat{\mathbf{u}}'|\mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned} \tag{1}
$$

With:

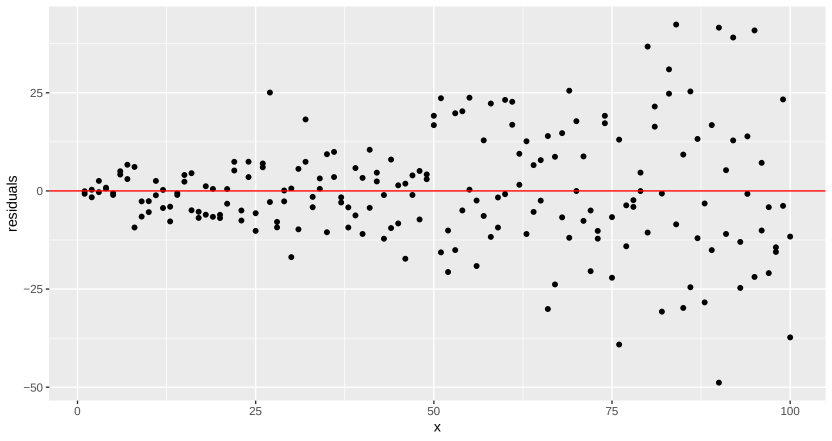
$$
E(\hat{\mathbf{u}}\hat{\mathbf{u}}'|\mathbf{X}) = \begin{bmatrix} \hat{\sigma}^2 & 0 & \dots & 0 \\ 0 & \hat{\sigma}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\sigma}^2 \end{bmatrix} = \hat{\sigma}^2\mathbf{I} = \Sigma
$$

Problem with heteroscedasticity: no constant error variance!

Heteroscedasticity: visualization



Heteroscedasticity: visualization



Heteroscedasticity

Let's start with fixing heteroscedasticity:

- Heteroscedasticity means that the variance of the disturbances, σ^2, in our sample is not constant.

Heteroscedasticity

Let's start with fixing heteroscedasticity:

- Heteroscedasticity means that the variance of the disturbances, σ^2, in our sample is not constant.
- Assumption 5, $Var(u|\mathbf{X}) = \sigma^2$, is violated and we cannot replace the population variance $Var(u|\mathbf{X})$ by $\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N-K-1}$

Heteroscedasticity

Let's start with fixing heteroscedasticity:

- Heteroscedasticity means that the variance of the disturbances, σ^2, in our sample is not constant.
- Assumption 5, $Var(u|\mathbf{X}) = \sigma^2$, is violated and we cannot replace the population variance $Var(u|\mathbf{X})$ by $\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N-K-1}$
- We can use **robust standard errors** to allow for such violation (while still assuming **no** serial correlation/autocorrelation)

Heteroscedasticity

Let's start with fixing heteroscedasticity:

- Heteroscedasticity means that the variance of the disturbances, σ^2, in our sample is not constant.
- Assumption 5, $Var(u|\mathbf{X}) = \sigma^2$, is violated and we cannot replace the population variance $Var(u|\mathbf{X})$ by $\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N-K-1}$
- We can use **robust standard errors** to allow for such violation (while still assuming **no** serial correlation/autocorrelation)
- We **estimate** the covariance matrix $\Sigma = E(\hat{\mathbf{u}}\hat{\mathbf{u}}'|\mathbf{X})$ using $\hat{\Sigma}$

$$\hat{\Sigma} = \begin{bmatrix} \hat{u}^2 & 0 & \dots & 0 \\ 0 & \hat{u}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{u}^2 \end{bmatrix}$$

White-robust estimators

- We use residuals of the prediction of y_i **by our model** for each i

White-robust estimators

- We use residuals of the prediction of y_i **by our model** for each i
- Purely empirical solution

White-robust estimators

- We use residuals of the prediction of y_i **by our model** for each i
- Purely empirical solution
- White has proved that the diagonal of matrix $\hat{\Sigma}$ is a consistent but biased estimator of $\Sigma = E(\hat{\mathbf{u}}\hat{\mathbf{u}}'|\mathbf{X})$ in small samples

White-robust estimators

- We use residuals of the prediction of y_i **by our model** for each i
- Purely empirical solution
- White has proved that the diagonal of matrix $\hat{\Sigma}$ is a consistent but biased estimator of $\Sigma = E(\hat{\mathbf{u}}\hat{\mathbf{u}}'|\mathbf{X})$ in small samples
- Your statistical software will multiply the squared residuals of each i by $n/df = n/(n-k-1)$ as a degree of freedom correction to have unbiased estimates

White-robust estimators

- We use residuals of the prediction of y_i **by our model** for each i
- Purely empirical solution
- White has proved that the diagonal of matrix $\hat{\Sigma}$ is a consistent but biased estimator of $\Sigma = E(\hat{\mathbf{u}}\hat{\mathbf{u}}'|\mathbf{X})$ in small samples
- Your statistical software will multiply the squared residuals of each i by $n/df = n/(n-k-1)$ as a degree of freedom correction to have unbiased estimates
- The estimate of the variance of the OLS estimator becomes

$$\widehat{Var}(\hat{\beta}) = \frac{n}{df}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

- This robust variance estimator is called **Sandwich estimator** giving us **White-Huber standard errors**

Robust estimation in R

Using R today we will:

1. Compute OLS and estimates of the S.E. using `lm()` and matrix algebra
2. Represent heteroscedasticity
3. Compute White-robust estimators for S.E. using `vcovHC()`, matrix algebra and `lm_robust()`
4. Perform a Breusch-Pagan test on heteroscedasticity

Robust estimation in `R`

Using `R` today we will:

1. Compute OLS and estimates of the S.E. using `lm()` and matrix algebra
2. Represent heteroscedasticity
3. Compute White-robust estimators for S.E. using `vcovHC()`, matrix algebra and `lm_robust()`
4. Perform a Breusch-Pagan test on heteroscedasticity

We are not doing it in Stata today, cause matrix algebra is definitely less handy... Equivalent commands exist, yet:

- To estimate robust standard errors: `reg ..., robust`
- To perform a Breusch-Pagan test: `estat hettest`
- To visualise heteroscedasticity after a regression: `rvfplot`

Serial correlation (time, space, both)

Now, serial correlation:

- Observations may be related within groups (e.g., within individuals, states, countries, within a unit across time periods). Typical problem with time series, spatial correlation, panel data

Serial correlation (time, space, both)

Now, serial correlation:

- Observations may be related within groups (e.g., within individuals, states, countries, within a unit across time periods). Typical problem with time series, spatial correlation, panel data
- With $i \neq j$,

$$E[u_{(i)g}, u_{(j)g'}] = \begin{cases} 0 & \text{if } g = g' \\ \sigma_{(ij)g} & \text{if } g \neq g' \end{cases}$$

Clustered observations

- Say, we have individuals clustered in G groups with N_g

Clustered observations

- Say, we have individuals clustered in G groups with N_g

$$Var(\mathbf{b}|\mathbf{X}) = E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right]$$

- with Σ given by

$$\begin{bmatrix}
\sigma^2_{(11)1} & \dots & \sigma^2_{(1N_1)1} & 0 & \dots & 0 & 0 & \dots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\sigma^2_{(N_11)1} & \dots & \sigma^2_{(N_1N_1)1} & 0 & \dots & 0 & 0 & \dots & 0 \\
0 & \dots & 0 & \sigma^2_{(11)2} & \dots & \sigma^2_{(1N_2)2} & 0 & \dots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \dots & 0 & \sigma^2_{(N_21)2} & \dots & \sigma^2_{(N_2N_2)1} & 0 & \dots & 0 \\
0 & \dots & 0 & 0 & \dots & 0 & \sigma^2_{(11)G} & \dots & \sigma^2_{(1N_G)G} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \dots & 0 & 0 & \dots & 0 & \sigma^2_{(N_G1)G} & \dots & \sigma^2_{(N_GN_G)G}
\end{bmatrix}$$

Dealing with clustered observations

What to do? We need an estimator for the variance that accounts for clustering:

- We practically never know $\sigma^2_{(N_g N_g)g}$: we need an estimate

Dealing with clustered observations

What to do? We need an estimator for the variance that accounts for clustering:

- We practically never know $\sigma^2_{(N_g N_g)g}$: we need an estimate
- We will use the residuals: the amount by which the prediction $\hat{\mathbf{y}} = \mathbf{X}\beta$ is off from the actual values of $\mathbf{y}$,

Dealing with clustered observations

What to do? We need an estimator for the variance that accounts for clustering:

- We practically never know $\sigma^2_{(N_g N_g)g}$: we need an estimate
- We will use the residuals: the amount by which the prediction $\hat{\mathbf{y}} = \mathbf{X}\beta$ is off from the actual values of $\mathbf{y}$, **but now** computed separated *within* each cluster of observations:

Dealing with clustered observations

What to do? We need an estimator for the variance that accounts for clustering:

- We practically never know $\sigma^2_{(N_g N_g)g}$: we need an estimate
- We will use the residuals: the amount by which the prediction $\hat{\mathbf{y}} = \mathbf{X}\beta$ is off from the actual values of $\mathbf{y}$, **but now** computed separated *within* each cluster of observations:

$$\frac{\sum_{i=1}^{N_g} \hat{u}_i^2}{N_g}, \quad i \in g \tag{2}$$

Dealing with clustered observations

What to do? We need an estimator for the variance that accounts for clustering:

- We practically never know $\sigma^2_{(N_g N_g)g}$: we need an estimate
- We will use the residuals: the amount by which the prediction $\hat{\mathbf{y}} = \mathbf{X}\beta$ is off from the actual values of $\mathbf{y}$, **but now** computed separated *within* each cluster of observations:

$$\frac{\sum_{i=1}^{N_g} \hat{u}_i^2}{N_g}, \quad i \in g \tag{2}$$

or $\hat{\mathbf{u}}_\mathbf{g}' \hat{\mathbf{u}}_\mathbf{g} / N_g$ where the numerator is SSR for all observations within each cluster.

Robust estimation and GLS

We can do even better:

- **Robust standard errors** use a robust covariance matrix. Keep the coefficient estimate but adjust **standard errors** to account for the violation of the assumption of i.i.d. $\rightarrow$ we get a conservative estimate!
- **Generalized least squares** proceed in two stages. First estimate the covariance matrix (stage 1) and then adjust coefficients (stage 2) $\rightarrow$ GLS may be more efficient

Generalized Least Squares (GLS)

- A family of estimators that generalise OLS procedures (OLS become only a particular case).

Generalized Least Squares (GLS)

- A family of estimators that generalise OLS procedures (OLS become only a particular case).
- Weighted Least Squares (WLS):
  $\hat{\beta} = (\mathbf{X}' \mathit{diag}\,[\mathbf{W}]\mathbf{X})^{-1}\mathbf{X}' \mathit{diag}\,[\mathbf{W}]\mathbf{Y}$

Generalized Least Squares (GLS)

- A family of estimators that generalise OLS procedures (OLS become only a particular case).
- Weighted Least Squares (WLS):
  $\hat{\beta} = (\mathbf{X}' \mathit{diag}\,[\mathbf{W}]\mathbf{X})^{-1}\mathbf{X}' \mathit{diag}\,[\mathbf{W}]\mathbf{Y}$
- You can use WLS with heteroscedasticity because you give a higher weight to data points with a smaller error variance, cause they are more informative!

Generalized Least Squares (GLS)

- A family of estimators that generalise OLS procedures (OLS become only a particular case).
- Weighted Least Squares (WLS):
  $\hat{\beta} = (\mathbf{X}' \mathit{diag}\,[\mathbf{W}]\mathbf{X})^{-1}\mathbf{X}' \mathit{diag}\,[\mathbf{W}]\mathbf{Y}$
- You can use WLS with heteroscedasticity because you give a higher weight to data points with a smaller error variance, cause they are more informative!
- You just need to specify the correct weight matrix **W**
- One common choice is to use the inverse of the OLS residuals (stage 1) as **W** for WLS (stage 2).

Generalized Least Squares (GLS)

- A family of estimators that generalise OLS procedures (OLS become only a particular case).
- Weighted Least Squares (WLS):
  $\hat{\beta} = (\mathbf{X}' \mathit{diag}\,[\mathbf{W}]\mathbf{X})^{-1}\mathbf{X}' \mathit{diag}\,[\mathbf{W}]\mathbf{Y}$
- You can use WLS with heteroscedasticity because you give a higher weight to data points with a smaller error variance, cause they are more informative!
- You just need to specify the correct weight matrix **W**
- One common choice is to use the inverse of the OLS residuals (stage 1) as **W** for WLS (stage 2). In this case we have Feasible Last Squares (FGLS)
- See an example in `R`

Take away points

1. **Always use robust standard errors!!!**
2. Consider whether your observations are clustered in a meaningful way, if yes, use clustered standard errors
3. Consider GLS as further robustness check on your estimation but (1) and (2) are mostly good enough.
4. R and Stata feature plenty of ways to implement robust and clustered standard errors; know how they work.

Non-parametric tests

Introduction

- There are alternatives to the t-test and F-test
- Choose your test based on your problem, and not the other way around
- Do not stick to parametric tests if your problem requires you to do otherwise

Pick an appropriate (non-parametric) test

1. Intro to non-parametric tests
2. Success probability: binomial test
3. Tests of differences between groups – paired from one-sample and independent from two-sample
    - Fisher sign test and Wilcoxon sign-rank test
    - Comparing success probabilities: Fisher's exact test

What are non-parametric tests

Non-parametric tests:

- Make fewer assumptions about the population – **distribution-free** tests

What are non-parametric tests

Non-parametric tests:

- Make fewer assumptions about the population – **distribution-free** tests
- Are more robust to outliers and heterogeneity of variance

What are non-parametric tests

Non-parametric tests:

- Make fewer assumptions about the population – **distribution-free** tests
- Are more robust to outliers and heterogeneity of variance
- Are robust even if you cannot reasonably characterize the distribution of the underlying population

What are non-parametric tests

Non-parametric tests:

- Make fewer assumptions about the population – **distribution-free** tests
- Are more robust to outliers and heterogeneity of variance
- Are robust even if you cannot reasonably characterize the distribution of the underlying population
- Are applicable for interval and ordinal data – some even for nominal data

What are non-parametric tests

Non-parametric tests:

- Make fewer assumptions about the population – **distribution-free** tests
- Are more robust to outliers and heterogeneity of variance
- Are robust even if you cannot reasonably characterize the distribution of the underlying population
- Are applicable for interval and ordinal data – some even for nominal data
- Have test statistics that are distributed normally when N is large

When are non-parametric tests advantageous

- When assumptions of parametric tests/estimators are not met. Example: t-statistic is not t-distributed if underlying population not normal or sample size too small

When are non-parametric tests advantageous

- When assumptions of parametric tests/estimators are not met. Example: t-statistic is not t-distributed if underlying population not normal or sample size too small
- Often easier to compute manually than their parametric alternatives

When are non-parametric tests advantageous

- When assumptions of parametric tests/estimators are not met. Example: t-statistic is not t-distributed if underlying population not normal or sample size too small
- Often easier to compute manually than their parametric alternatives
- We can often derive the exact distribution of the test statistic

Disadvantages of non-parametric tests

- Mostly no estimates of variance

Disadvantages of non-parametric tests

- Mostly no estimates of variance
- Mostly no confidence intervals

Disadvantages of non-parametric tests

- Mostly no estimates of variance
- Mostly no confidence intervals
- Need more observations to draw a conclusion with same certainty

Disadvantages of non-parametric tests

- Mostly no estimates of variance
- Mostly no confidence intervals
- Need more observations to draw a conclusion with same certainty i.e., less powerful as parametric alternative when assumptions for parametric tests are met

Disadvantages of non-parametric tests

- Mostly no estimates of variance
- Mostly no confidence intervals
- Need more observations to draw a conclusion with same certainty i.e., less powerful as parametric alternative when assumptions for parametric tests are met
- **Yet** often differences are small, and parametric alternatives perform vastly worse when assumptions are not met

Binomial test: Basics

- Observing outcome B of n independent repeated Bernoulli trials, can I say that success probability for each trial $p = p_0$? $P(p_0|B, n)$?

Binomial test: Basics

- Observing outcome B of n independent repeated Bernoulli trials, can I say that success probability for each trial $p = p_0$? $P(p_0|B, n)$? Example: you toss 5 times a coin and count the number of Heads. You get Heads $= 3$. Can we say $p = .5$?

Binomial test: Basics

- Observing outcome B of n independent repeated Bernoulli trials, can I say that success probability for each trial $p = p_0$? $P(p_0|B, n)$? Example: you toss 5 times a coin and count the number of Heads. You get Heads = 3. Can we say $p = .5$?
- Assumptions:
  1. Dichotomous data: outcomes are success or failure
  2. p remains constant for each trial
  3. n trials are independent
- $H_0 : p = p_0$

Binomial test: Basics

- Observing outcome B of n independent repeated Bernoulli trials, can I say that success probability for each trial $p = p_0$? $P(p_0|B, n)$? Example: you toss 5 times a coin and count the number of Heads. You get Heads = 3. Can we say $p = .5$?
- Assumptions:
  1. Dichotomous data: outcomes are success or failure
  2. p remains constant for each trial
  3. n trials are independent
- $H_0 : p = p_0$
- Other tests/statistics below relate to the basic binomial test of significance

Binomial test: Basics

- Observing outcome B of n independent repeated Bernoulli trials, can I say that success probability for each trial $p = p_0$? $P(p_0|B, n)$? Example: you toss 5 times a coin and count the number of Heads. You get Heads = 3. Can we say $p = .5$?
- Assumptions:
  1. Dichotomous data: outcomes are success or failure
  2. p remains constant for each trial
  3. n trials are independent
- $H_0 : p = p_0$
- Other tests/statistics below relate to the basic binomial test of significance
- Under assumptions 1-3, it is a distribution-free test of H_0 because the probability distribution of B is determined without further assumptions on the distribution of the underlying population

Binomial test: Procedure

- To test $H_0 : p = p_0$ we set the desired level of significance α and we set B to number of observed successes with n trials

Binomial test: Procedure

- To test $H_0 : p = p_0$ we set the desired level of significance α and we set B to number of observed successes with n trials
- Build the binomial distribution under the null-hypothesis

Binomial test: Procedure

- To test $H_0 : p = p_0$ we set the desired level of significance α and we set B to number of observed successes with n trials
- Build the binomial distribution under the null-hypothesis
- Reject H_0 if $B \geq b_{\alpha_2}$ or $B \leq b_{\alpha_1}$

Binomial test: Procedure

- To test $H_0 : p = p_0$ we set the desired level of significance α and we set B to number of observed successes with n trials
- Build the binomial distribution under the null-hypothesis
- Reject H_0 if $B \geq b_{\alpha_2}$ or $B \leq b_{\alpha_1}$
- Where b_{α_2} is the upper α_2 percentile point and b_{α_1} is the lower α_1 percentile point with $\alpha = \alpha_1 + \alpha_2$

Binomial test: Example

- Say $n = 8$ and we test $H_0 : p = .4$ vs $H_1 : p \neq .4$

Binomial test: Example

- Say $n = 8$ and we test $H_0 : p = .4$ vs $H_1 : p \neq .4$
- You can use the table of the binomial distribution or your statistical software and quantile function to obtain the binomial distribution for $n = 8$ and $p = .4$. Suppose that $\alpha = .05$. Under the null-hypothesis the distribution:

Binomial test: Example

- Say $n = 8$ and we test $H_0 : p = .4$ vs $H_1 : p \neq .4$
- You can use the table of the binomial distribution or your statistical software and quantile function to obtain the binomial distribution for $n = 8$ and $p = .4$. Suppose that $\alpha = .05$. Under the null-hypothesis the distribution:
  - leaves 2.5% of the observations to the left of 1
  - leaves 2.5% of the observations to the right of 6
- We reject $H_0 : p = .4$ if 6 or more successes are observed as well as if 1 or less successes are observed

Sign test (Fisher): Basics

- Alternative to paired t-test (*e.g.* difference-in-means) which assumes normality and equal variance across groups in underlying data

Sign test (Fisher): Basics

- Alternative to paired t-test (*e.g.* difference-in-means) which assumes normality and equal variance across groups in underlying data
- Information taken from **signs** in difference between paired observations

Sign test (Fisher): Basics

- Alternative to paired t-test (*e.g.* difference-in-means) which assumes normality and equal variance across groups in underlying data
- Information taken from **signs** in difference between paired observations
- Assumptions:

Sign test (Fisher): Basics

- Alternative to paired t-test (*e.g.* difference-in-means) which assumes normality and equal variance across groups in underlying data
- Information taken from **signs** in difference between paired observations
- Assumptions:
  - paired observations $(X_1^1, X_1^2), \ldots, (X_N^1, X_N^2)$ are **random** sample and iid
  - paired observations are dependent
  - paired differences come from same continuous distribution

Sign test (Fisher): Basics

- Alternative to paired t-test (*e.g.* difference-in-means) which assumes normality and equal variance across groups in underlying data
- Information taken from **signs** in difference between paired observations
- Assumptions:
  - paired observations $(X_1^1, X_1^2), \ldots, (X_N^1, X_N^2)$ are **random** sample and iid
  - paired observations are dependent
  - paired differences come from same continuous distribution
- Use when **direction** of difference between two measurements on same unit can be determined

Sign test: Procedure

- Compute differences $D_i = X_i^1 - X_i^2$ between N pairs of matched observations.
- Count the number of positive and negative D_i, n^+ and n^-. **We do not care about their size!**

Sign test: Procedure

- Compute differences $D_i = X_i^1 - X_i^2$ between N pairs of matched observations.
- Count the number of positive and negative D_i, n^+ and n^-. **We do not care about their size!**
- If there is no difference in the samples, then the median of the distribution of D_i, θ, should be 0. $H_0 : \theta = 0$ vs $H_1 : \theta > 0$.

Sign test: Procedure

- Compute differences $D_i = X_i^1 - X_i^2$ between N pairs of matched observations.
- Count the number of positive and negative D_i, n^+ and n^-. **We do not care about their size!**
- If there is no difference in the samples, then the median of the distribution of D_i, θ, should be 0. $H_0 : \theta = 0$ vs $H_1 : \theta > 0$.
- n^+ is our test statistic. What is its distribution? It's binomial if we consider "being positive" as a Bernoulli random variable!

Sign test: Procedure

- Compute differences $D_i = X_i^1 - X_i^2$ between N pairs of matched observations.
- Count the number of positive and negative D_i, n^+ and n^-. **We do not care about their size!**
- If there is no difference in the samples, then the median of the distribution of D_i, θ, should be 0. $H_0 : \theta = 0$ vs $H_1 : \theta > 0$.
- n^+ is our test statistic. What is its distribution? It's binomial if we consider "being positive" as a Bernoulli random variable!
- Calculate the binomial coefficient $B = \binom{N}{n^+}$. $B/2^N$ is the probability of getting **exactly** the amount of n^+ and n^- under the null.
- To get a p-value, sum all binomial coefficients that are $\leq B$ and divide by 2^N

Sign test: Example

- Consider the table below. (`income`-variable in `gssData.dta`)

Sign test: Example

- Consider the table below. (`income`-variable in `gssData.dta`)
- Question: Did income increase from '08 to '12?
- Note, it is an ordinal measured variable, taking a difference may not make sense – assume for this example that it does

Sign test: Example

- Consider the table below. (`income`-variable in `gssData.dta`)
- Question: Did income increase from '08 to '12?
- Note, it is an ordinal measured variable, taking a difference may not make sense – assume for this example that it does

```
    +-----------------------------+
    | income08   income12    D_i |
    |-----------------------------|
 1. |        3         19    -16 |
 2. |        4          9     -5 |
 3. |       17         20     -3 |
 4. |       15         17     -2 |
 5. |       14         16     -2 |
...
```

Sign test: Example

- Differences:

Sign test: Example

- Differences: -10 -9 -8 -6 -4 -4 -3 0 **1 1 2 2 3 5 16**

Sign test: Example

- Differences: -10 -9 -8 -6 -4 -4 -3 0 **1 1 2 2 3 5 16**
- How likely is it to observe 7 positive D_i when $H_0 : \theta = 0$ is true, thus with no difference between income in 2008 and 2012?

Sign test: Example

- Differences: -10 -9 -8 -6 -4 -4 -3 0 **1 1 2 2 3 5 16**
- How likely is it to observe 7 positive D_i when $H_0 : \theta = 0$ is true, thus with no difference between income in 2008 and 2012?

$$B = \binom{N}{n^+} = \frac{15!}{8!7!} = 6435$$

- Thus we have that the probability of getting exactly 7 positive differences is $B/2^N = 6435/2^{15} = 0.19$
- To get the related p-value we should sum all binomial coefficients that are $\leq B$ and divide the sum by 2^N.

Sign test: Example

- Differences: -10 -9 -8 -6 -4 -4 -3 0 **1 1 2 2 3 5 16**
- How likely is it to observe 7 positive D_i when $H_0 : \theta = 0$ is true, thus with no difference between income in 2008 and 2012?

$$B = \binom{N}{n^+} = \frac{15!}{8!7!} = 6435$$

- Thus we have that the probability of getting exactly 7 positive differences is $B/2^N = 6435/2^{15} = 0.19$
- To get the related p-value we should sum all binomial coefficients that are $\leq B$ and divide the sum by 2^N.
- We can do quicker: What is the distribution of D_i?

Sign test: Example

- Differences: -10 -9 -8 -6 -4 -4 -3 0 **1 1 2 2 3 5 16**
- How likely is it to observe 7 positive D_i when $H_0 : \theta = 0$ is true, thus with no difference between income in 2008 and 2012?

$$B = \binom{N}{n^+} = \frac{15!}{8!7!} = 6435$$

- Thus we have that the probability of getting exactly 7 positive differences is $B/2^N = 6435/2^{15} = 0.19$
- To get the related p-value we should sum all binomial coefficients that are $\leq B$ and divide the sum by 2^N.
- We can do quicker: What is the distribution of D_i?
- Binomial with $N = 15$, $p = .5$ (under H_0), and $x = 7$.

Sign test: Example

- Differences: -10 -9 -8 -6 -4 -4 -3 0 **1 1 2 2 3 5 16**
- How likely is it to observe 7 positive D_i when $H_0 : \theta = 0$ is true, thus with no difference between income in 2008 and 2012?

$$B = \binom{N}{n^+} = \frac{15!}{8!7!} = 6435$$

- Thus we have that the probability of getting exactly 7 positive differences is $B/2^N = 6435/2^{15} = 0.19$
- To get the related p-value we should sum all binomial coefficients that are $\leq B$ and divide the sum by 2^N.
- We can do quicker: What is the distribution of D_i?
- Binomial with $N = 15$, $p = .5$ (under H_0), and $x = 7$.
- Our p-value is $Prob(X \leq 7) = 0.50$

Sign test: Example

- Differences: -10 -9 -8 -6 -4 -4 -3 0 **1 1 2 2 3 5 16**
- How likely is it to observe 7 positive D_i when $H_0 : \theta = 0$ is true, thus with no difference between income in 2008 and 2012?

$$B = \binom{N}{n^+} = \frac{15!}{8!7!} = 6435$$

- Thus we have that the probability of getting exactly 7 positive differences is $B/2^N = 6435/2^{15} = 0.19$
- To get the related p-value we should sum all binomial coefficients that are $\leq B$ and divide the sum by 2^N.
- We can do quicker: What is the distribution of D_i?
- Binomial with $N = 15$, $p = .5$ (under H_0), and $x = 7$.
- Our p-value is $Prob(X \leq 7) = 0.50$
- Statistical packages will then compute these for you

Sign test: Small sample issues

- Only needs **sign** of difference, not difference itself

Sign test: Small sample issues

- Only needs **sign** of difference, not difference itself
- Less efficient than Wilcoxon sign-rank test (uses only sign not ordering) but rather robust to outliers

Sign test: Small sample issues

- Only needs **sign** of difference, not difference itself
- Less efficient than Wilcoxon sign-rank test (uses only sign not ordering) but rather robust to outliers
- How to deal with $D_i = 0$? Usually ignored but reduces effective sample size

Sign test: Small sample issues

- Only needs **sign** of difference, not difference itself
- Less efficient than Wilcoxon sign-rank test (uses only sign not ordering) but rather robust to outliers
- How to deal with $D_i = 0$? Usually ignored but reduces effective sample size
- Works for interval data but pay attention to ties - which correction for tied values?

Wilcoxon sign-rank test: Basics

- Alternative to paired t-test which assumes normality and equal variance across groups in underlying data

Wilcoxon sign-rank test: Basics

- Alternative to paired t-test which assumes normality and equal variance across groups in underlying data
- Information taken from signs in difference between paired observations (pre- vs post-treatment)

Wilcoxon sign-rank test: Basics

- Alternative to paired t-test which assumes normality and equal variance across groups in underlying data
- Information taken from signs in difference between paired observations (pre- vs post-treatment)
- When actual difference pre- vs post-treatment is greater than 0, tendency to larger proportion of positive differences

Wilcoxon sign-rank test: Basics

- Assumptions:

Wilcoxon sign-rank test: Basics

- Assumptions:
  - paired observations $(X_1^1, X_1^2), \ldots, (X_N^1, X_N^2)$ are random sample and iid

Wilcoxon sign-rank test: Basics

- Assumptions:
  - paired observations $(X_1^1, X_1^2), \ldots, (X_N^1, X_N^2)$ are random sample and iid i.e., differences are mutually independent (while paired observations are dependent)

Wilcoxon sign-rank test: Basics

- Assumptions:
  - paired observations $(X_1^1, X_1^2), \ldots, (X_N^1, X_N^2)$ are random sample and iid i.e., differences are mutually independent (while paired observations are dependent)
  - paired differences come from a continuous distribution

Wilcoxon sign-rank test: Basics

- Assumptions:
  - paired observations $(X_1^1, X_1^2), \ldots, (X_N^1, X_N^2)$ are random sample and iid i.e., differences are mutually independent (while paired observations are dependent)
  - paired differences come from a continuous distribution
- Test of null hypothesis of zero shift in location (no treatment effect), $H_0 : \theta = 0$ – null hypothesis states that each of the distributions for the differences is symmetrically distributed around the 0

Wilcoxon sign-rank test: Basics

- Assumptions:
  - paired observations $(X_1^1, X_1^2), \ldots, (X_N^1, X_N^2)$ are random sample and iid i.e., differences are mutually independent (while paired observations are dependent)
  - paired differences come from a continuous distribution
- Test of null hypothesis of zero shift in location (no treatment effect), $H_0 : \theta = 0$ – null hypothesis states that each of the distributions for the differences is symmetrically distributed around the 0
- Use when direction of difference **and magnitude** between two measurements on same unit can be determined

Wilcoxon sign-rank test: Procedure

- Compute difference $D_i = X_i^1 - X_i^2$ between N pairs of matched observations
- Order **absolute values** of differences from smallest to largest

Wilcoxon sign-rank test: Procedure

- Compute difference $D_i = X_i^1 - X_i^2$ between N pairs of matched observations
- Order **absolute values** of differences from smallest to largest
- Let S_i denote the rank of D_i in the joint ordering (assign average rank to ties)

Wilcoxon sign-rank test: Procedure

- Compute difference $D_i = X_i^1 - X_i^2$ between N pairs of matched observations
- Order **absolute values** of differences from smallest to largest
- Let S_i denote the rank of D_i in the joint ordering (assign average rank to ties)
- "Wilcoxon signed rank statistic", W^+, is sum of positive signed ranks
- Under $H_0 : \theta = 0$, W^+ is distributed according to the distribution derived by Wilcoxon (1954)
- Reject H_0 if $W^+ \geq w_{\alpha/2}$ or $W^+ \leq \frac{N(N+2)}{2} - w_{\alpha/2}$
- distribution of W is based on permutations of all possible rankings

Wilcoxon sign-rank test: Example

$N = 15$

```
    +-------------------------------------------+
    | income08   income12   D_i   absD_i   rank |
    |-------------------------------------------|
 1. |       22         22     0        0      1 |
 2. |       20         21    -1        1    2.5 |
 3. |       20         21    -1        1    2.5 |
 4. |       14         16    -2        2    4.5 |
 5. |       15         17    -2        2    4.5 |
    |-------------------------------------------|
 6. |       16         13     3        3    6.5 |
 7. |       17         20    -3        3    6.5 |
 8. |       21         17     4        4    8.5 |
...
```

Wilcoxon sign-rank test: Distribution of W^+

- Sum of ranks of positive differences: 74.5 (Sum of ranks of negative difference: 45.5)

Wilcoxon sign-rank test: Distribution of W^+

- Sum of ranks of positive differences: 74.5 (Sum of ranks of negative difference: 45.5)
- Lowest possible sum of ranks of positive differences?

Wilcoxon sign-rank test: Distribution of W^+

- Sum of ranks of positive differences: 74.5 (Sum of ranks of negative difference: 45.5)
- Lowest possible sum of ranks of positive differences? 0 – no absolute value of differences is positive

Wilcoxon sign-rank test: Distribution of W^+

- Sum of ranks of positive differences: 74.5 (Sum of ranks of negative difference: 45.5)
- Lowest possible sum of ranks of positive differences? 0 – no absolute value of differences is positive
- Highest possible? $N(N+1)/2 = 15(16)/2 = 120$ – all differences are positive

Wilcoxon sign-rank test: Distribution of W^+

- Sum of ranks of positive differences: 74.5 (Sum of ranks of negative difference: 45.5)
- Lowest possible sum of ranks of positive differences? 0 – no absolute value of differences is positive
- Highest possible? $N(N+1)/2 = 15(16)/2 = 120$ – all differences are positive
- What is the smallest significance level at which these data lead to rejection of H_0?

Wilcoxon sign-rank test: Distribution of W^+

- Sum of ranks of positive differences: 74.5 (Sum of ranks of negative difference: 45.5)
- Lowest possible sum of ranks of positive differences? 0 – no absolute value of differences is positive
- Highest possible? $N(N+1)/2 = 15(16)/2 = 120$ – all differences are positive
- What is the smallest significance level at which these data lead to rejection of H_0?
- Consulting statistical tables of W^+ distribution, for our example we have $N = 14$ and $W^+ = 74.5$

.117	$n = 14$	66	.213		93
.102		67	.196		94
.088		68	.179		95
.076		69	.163		96
.065		70	.148		97
.055		71	.134		98
.046		72	.121		99
.039		73	.108		100
.032		74	.097		101
.026		75	.086		102
.021		76	.077		103
.017		77	.068		104
.013		78	.059		105
.010		79	.052		106
.008		80	.045		107
.006		81	.039		108
.005		82	.034		109
.003		83	.029		110
.002		84	.025		111
.002		85	.021		112
.001		86	.018		113
.001		87	.015		114
<.0005		88	.012	$n = 16$	93
.207		89	.010		94
.188		90	.008		95
.170		91	.007		96
.153		92	.005		97
.137		93	.004		98
.122		94	.003		99
.108		95	.003		100
.095		96	.002		101
.084		97	.002		102
.073		98	.001		103
.064		99	.001		104
.055		100	.001		105
.047		101	<.0005		106
					107

Wilcoxon sign-rank test: Example

- Reject H_0 if $W^+ \geq w_{\alpha/2}$ or $W^+ \leq \frac{n(n+2)}{2} - w_{\alpha/2}$

Wilcoxon sign-rank test: Example

- Reject H_0 if $W^+ \geq w_{\alpha/2}$ or $W^+ \leq \frac{n(n+2)}{2} - w_{\alpha/2}$
- For $\alpha = .052$, reject if $W^+ \geq 80$ or $W^+ \leq \frac{15(16)}{2} - 80 = 120 - 80 = 40$

Wilcoxon sign-rank test: Example

- Reject H_0 if $W^+ \geq w_{\alpha/2}$ or $W^+ \leq \frac{n(n+2)}{2} - w_{\alpha/2}$
- For $\alpha = .052$, reject if $W^+ \geq 80$ or $W^+ \leq \frac{15(16)}{2} - 80 = 120 - 80 = 40$ – so we do not reject

Wilcoxon sign-rank test: Example

- Reject H_0 if $W^+ \geq w_{\alpha/2}$ or $W^+ \leq \frac{n(n+2)}{2} - w_{\alpha/2}$
- For $\alpha = .052$, reject if $W^+ \geq 80$ or $W^+ \leq \frac{15(16)}{2} - 80 = 120 - 80 = 40$ – so we do not reject
- Large sample approximation:

Wilcoxon sign-rank test: Example

- Reject H_0 if $W^+ \geq w_{\alpha/2}$ or $W^+ \leq \frac{n(n+2)}{2} - w_{\alpha/2}$
- For $\alpha = .052$, reject if $W^+ \geq 80$ or $W^+ \leq \frac{15(16)}{2} - 80 = 120 - 80 = 40$ – so we do not reject
- Large sample approximation:
  - Standardize W^+: Under the null, $E(W^+) = \frac{n(n+1)}{4}$ and $var(W^+) = \frac{n(n+1)(2n+1)}{24}$

Wilcoxon sign-rank test: Example

- Reject H_0 if $W^+ \geq w_{\alpha/2}$ or $W^+ \leq \frac{n(n+2)}{2} - w_{\alpha/2}$
- For $\alpha = .052$, reject if $W^+ \geq 80$ or $W^+ \leq \frac{15(16)}{2} - 80 = 120 - 80 = 40$ – so we do not reject
- Large sample approximation:
  - Standardize W^+: Under the null, $E(W^+) = \frac{n(n+1)}{4}$ and $var(W^+) = \frac{n(n+1)(2n+1)}{24}$
  - $W^* = \frac{W^+ - E(W^+)}{\sqrt{var(W^+)}}$

Wilcoxon sign-rank test: Example

- Reject H_0 if $W^+ \geq w_{\alpha/2}$ or $W^+ \leq \frac{n(n+2)}{2} - w_{\alpha/2}$
- For $\alpha = .052$, reject if $W^+ \geq 80$ or $W^+ \leq \frac{15(16)}{2} - 80 = 120 - 80 = 40$ – so we do not reject
- Large sample approximation:
  - Standardize W^+: Under the null, $E(W^+) = \frac{n(n+1)}{4}$ and $var(W^+) = \frac{n(n+1)(2n+1)}{24}$
  - $W^* = \frac{W^+ - E(W^+)}{\sqrt{var(W^+)}}$
  - With $n \to \infty$, $W^+ \sim N(0,1)$

Wilcoxon sign-rank test: Example

- Reject H_0 if $W^+ \geq w_{\alpha/2}$ or $W^+ \leq \frac{n(n+2)}{2} - w_{\alpha/2}$
- For $\alpha = .052$, reject if $W^+ \geq 80$ or $W^+ \leq \frac{15(16)}{2} - 80 = 120 - 80 = 40$ – so we do not reject
- Large sample approximation:
  - Standardize W^+: Under the null, $E(W^+) = \frac{n(n+1)}{4}$ and $var(W^+) = \frac{n(n+1)(2n+1)}{24}$
  - $W^* = \frac{W^+ - E(W^+)}{\sqrt{var(W^+)}}$
  - With $n \to \infty$, $W^+ \sim N(0, 1)$
  - Adjustments for ties are needed
- Your statistical package will perform this test for you, but this is what happens under the surface

Wilcoxon sign-rank test: Small sample issues

- Wilcoxon sign-rank test incorporates more information than Fisher's sign-test but also needs more information

Wilcoxon sign-rank test: Small sample issues

- Wilcoxon sign-rank test incorporates more information than Fisher's sign-test but also needs more information
- How to deal with $D_i = 0$? Usually ignored but reduces effective sample size

Wilcoxon sign-rank test: Small sample issues

- Wilcoxon sign-rank test incorporates more information than Fisher's sign-test but also needs more information
- How to deal with $D_i = 0$? Usually ignored but reduces effective sample size
- Works for interval data but pay attention to ties - which correction for tied values?

Fisher's exact test – Intuition

When do we use it?

1. When we have two independent samples arranged in a contingency table
2. When in each sample we measure the frequency of success *vs* failure (Bernoulli)
3. When expected frequency of any success or failure is below 5. Otherwise use χ^2 test

Fisher's exact test – Three cases

1. Row totals are fixed, column totals are random (or the other way around)

Fisher's exact test – Three cases

1. Row totals are fixed, column totals are random (or the other way around)
   - The hypothesis tested is $p_1 = p_2$

Fisher's exact test – Three cases

1. Row totals are fixed, column totals are random (or the other way around)
    - The hypothesis tested is $p_1 = p_2$
2. Both row totals and column totals are random

Fisher's exact test – Three cases

1. Row totals are fixed, column totals are random (or the other way around)
   - The hypothesis tested is $p_1 = p_2$
2. Both row totals and column totals are random
   - The hypothesis tested is bivariate independence

Fisher's exact test – Three cases

1. Row totals are fixed, column totals are random (or the other way around)
    - The hypothesis tested is $p_1 = p_2$
2. Both row totals and column totals are random
    - The hypothesis tested is bivariate independence
3. Both row totals and column totals are fixed

Fisher's exact test – Three cases

1. Row totals are fixed, column totals are random (or the other way around)
    - The hypothesis tested is $p_1 = p_2$
2. Both row totals and column totals are random
    - The hypothesis tested is bivariate independence
3. Both row totals and column totals are fixed
    - The hypothesis tested is independence

Fisher's exact test – Basics

1. indicate as O_{1S} the *S*uccessful outcome you observe from sample 1 (similarly you have O_{1F}, O_{2S}, O_{2F}).
2. Indicate as $n_{1\cdot}$ ($n_{2\cdot}$) the independent repeated Bernoulli trials from sample 1 (2), each one with success probability p_1 (p_2)

We will have:

	Successes	Failures	Totals
Sample 1	O_{1S}	O_{1F}	$n_{1\cdot}$
Sample 2	O_{2S}	O_{2F}	$n_{2\cdot}$
Totals	$n_{\cdot S}$	$n_{\cdot F}$	n

$$H_0 : p_1 = p_2 = p$$
$$H_1 : p_1 \neq p_2$$

Fisher's exact test – Procedure

- In effect the test asks "what is the probability of having a table as extreme as the one we observe, if the null hypothesis is true"?
- Hypergeometric distribution: $Prob(O_{1S} = x | n_{1\cdot}, n_{2\cdot}, n_{\cdot S}, n_{\cdot F}) = \frac{n_{1\cdot}! n_{2\cdot}! n_{\cdot S}! n_{\cdot F}!}{n! x! O_{1F}! O_{2S}! O_{2F}!}$
- Fisher's exact test rejects $H_0 : p_1 = p_2$ if your observed $O_{1S} \geq q_\alpha$
- Where q_α is chosen from the conditional distribution described above so that $Prob(O_{1S} = q_\alpha | n_{1\cdot}, n_{2\cdot}, n_{\cdot S}, n_{\cdot F}) = \alpha$ where α is our desired level of significance

Fisher's exact test – Examples

Suppose you are curing two independent samples of patients, one gets an innovative drug and the other gets the traditional one. You measure success as cure, failure as not cure

	S (cured)	F (not cured)	Totals
S1 (new drug)	5	4	9
S2 (old drug)	4	2	6
Totals	9	6	15

- Assume horizontal and vertical totals are fixed
- $H_0 : p_1 < p_2$
- Under the null-hypothesis, what are the probabilities of the tables that would give us a value as small as or smaller than the observed value of $O_{1S} = 5$?

Fisher's exact test – Example

Sample 1	3	6	4	5	5	4	6	3	7	2	8	1	9	0
Sample 2	6	0	5	1	4	2	3	3	2	4	1	5	0	6
p	.017		.151		.378		.336		.108		.011		.000	

- $H_0 : p_1 < p_2$
- Under the null-hypothesis, what are the probabilities of the tables that would give us a value as small as or smaller than the observed value of $O_{1S} = 5$?
- It will be $p = .017 + .151 + .378 = .546$

Fisher's exact test – Example

If (with the same horizontal and vertical totals) what we observe were:

	S (cured)	F (not cured)	Totals
S1 (new drug)	8	1	9
S2 (old drug)	1	5	6
Totals	9	6	15

The probability of having a table as extreme (or more) as the one you observe would be: $p = 0.0107 + 0.0002 = 0.0109$!

Conclusion

All clear? Questions?

Thanks and see you next week!