

Research Design

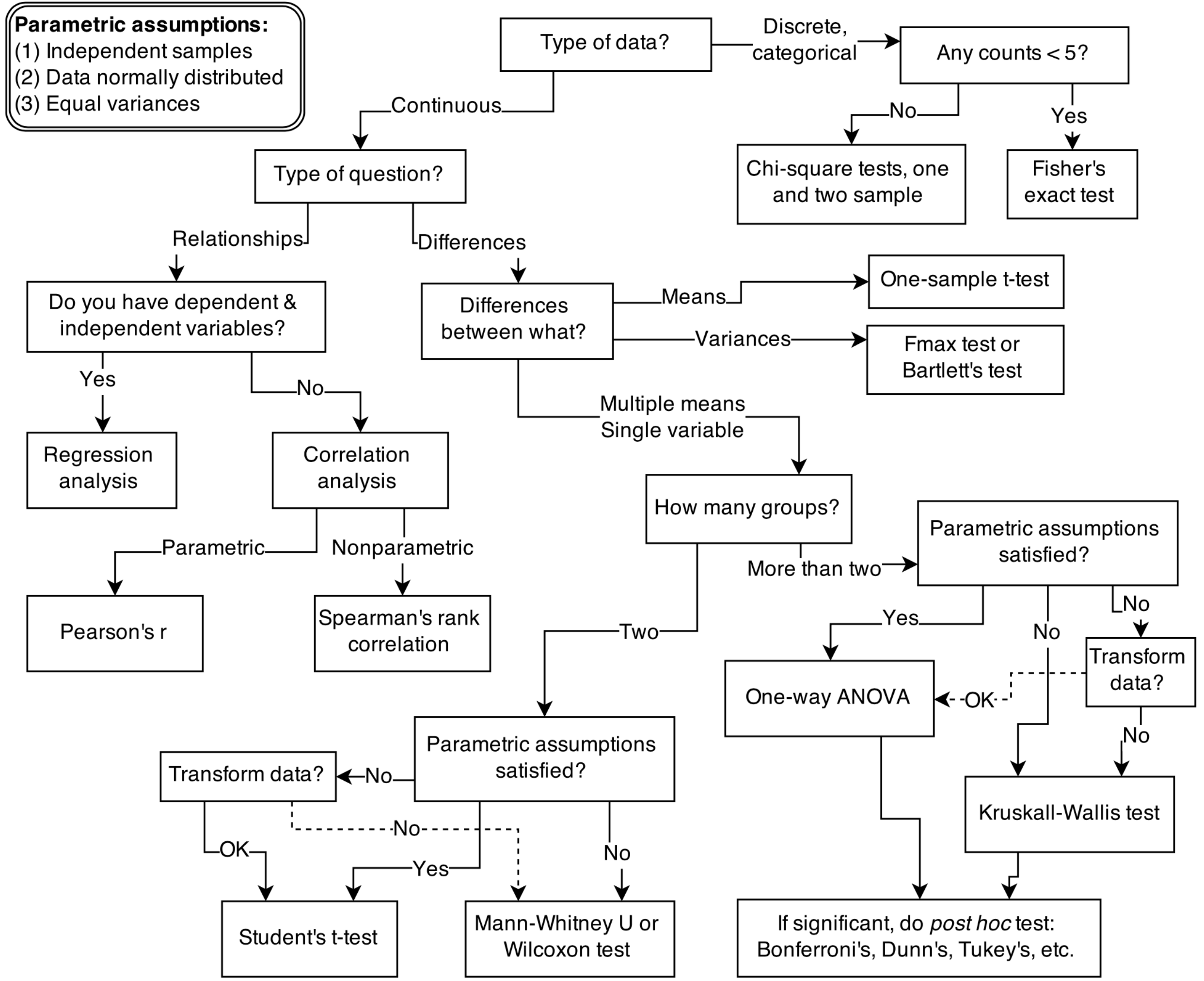
05: statistical tests continued; inference and interpretation

Initial questions about the pre-lecture notes?

‘a zoo of pre-constructed golems known as “tests” ’

— Richard McElreath

a zoo of tests, a decision tree for selecting one (and *not all* named or unnamed tests are listed below)

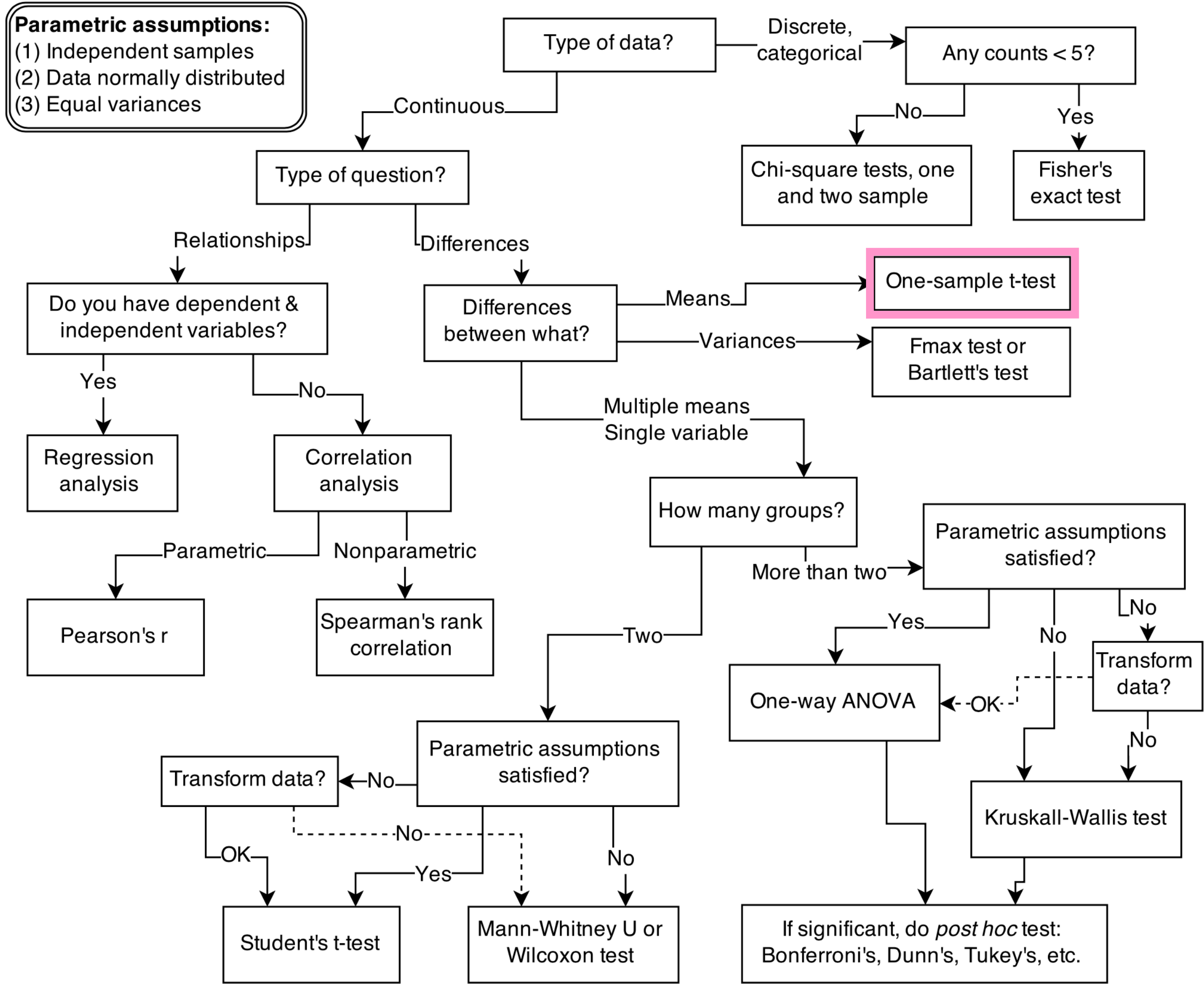


This zoo of tests does share common theories based on probability.

— Casella & Berger 1990; Lehmann & Casella 1998; Lehmann & Romano 2005

zoo & decisions, comparing \bar{x} to μ , $x \in \mathbb{R}$

Recall Student's t test — comparing one sample means to μ where we assume a population normal distribution with unknown standard deviation σ .



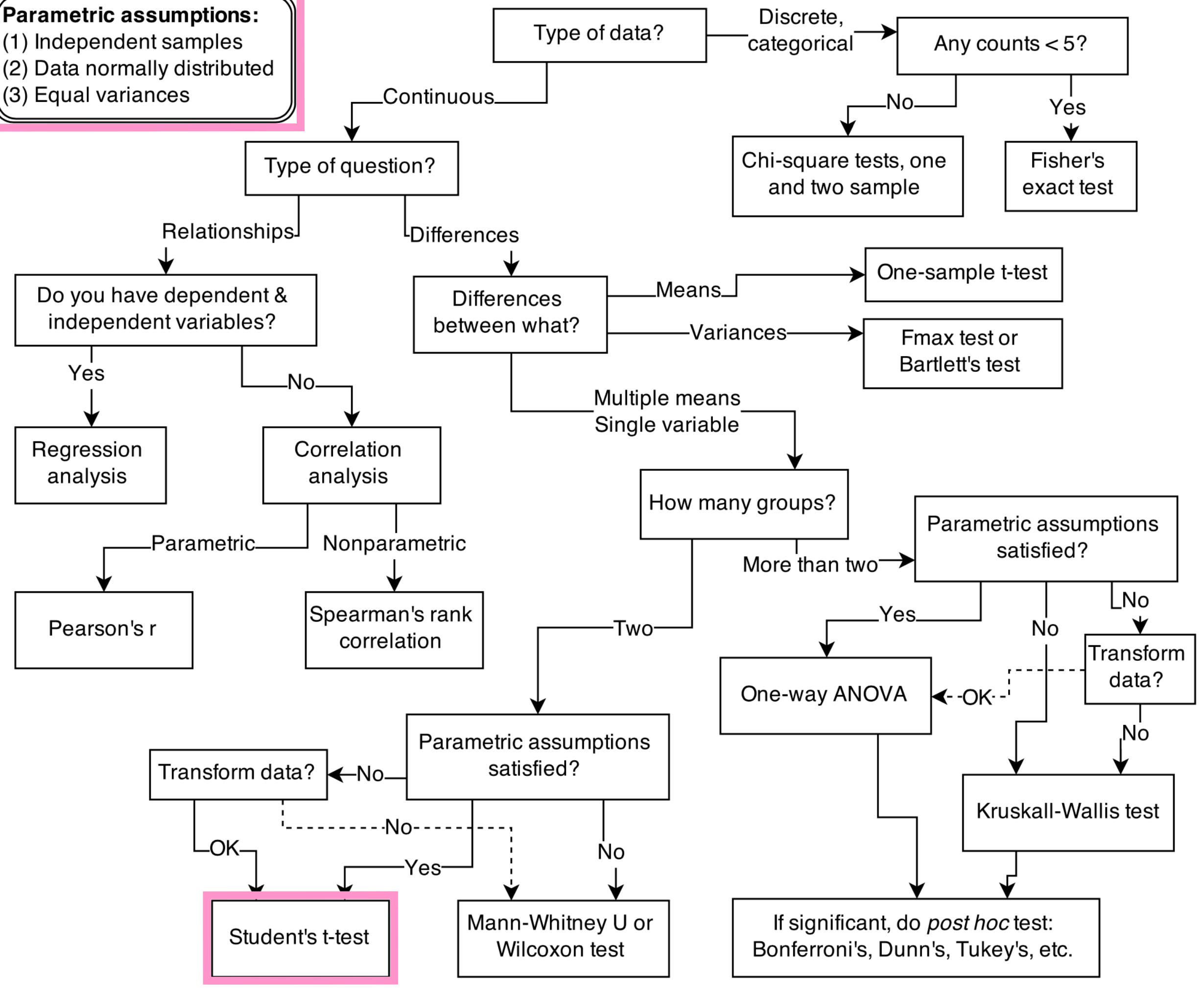
$$H_0 : \mu = 0, H_A : \mu \neq 0$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}, \quad \nu = n - 1, \quad p = F_T(t, \nu)$$

```
t.test(x, mu, alternative = "two.sided", conf.level = 0.95)
```

zoo & decisions, comparing locations, data as \mathbb{R}

Parametric assumptions:
 (1) Independent samples
 (2) Data normally distributed
 (3) Equal variances



Student's t test — comparing two sample means where we can assume an underlying normal distribution.

$$H_0 : \mu_1 = \mu_2, H_A : \mu_1 \neq \mu_2$$

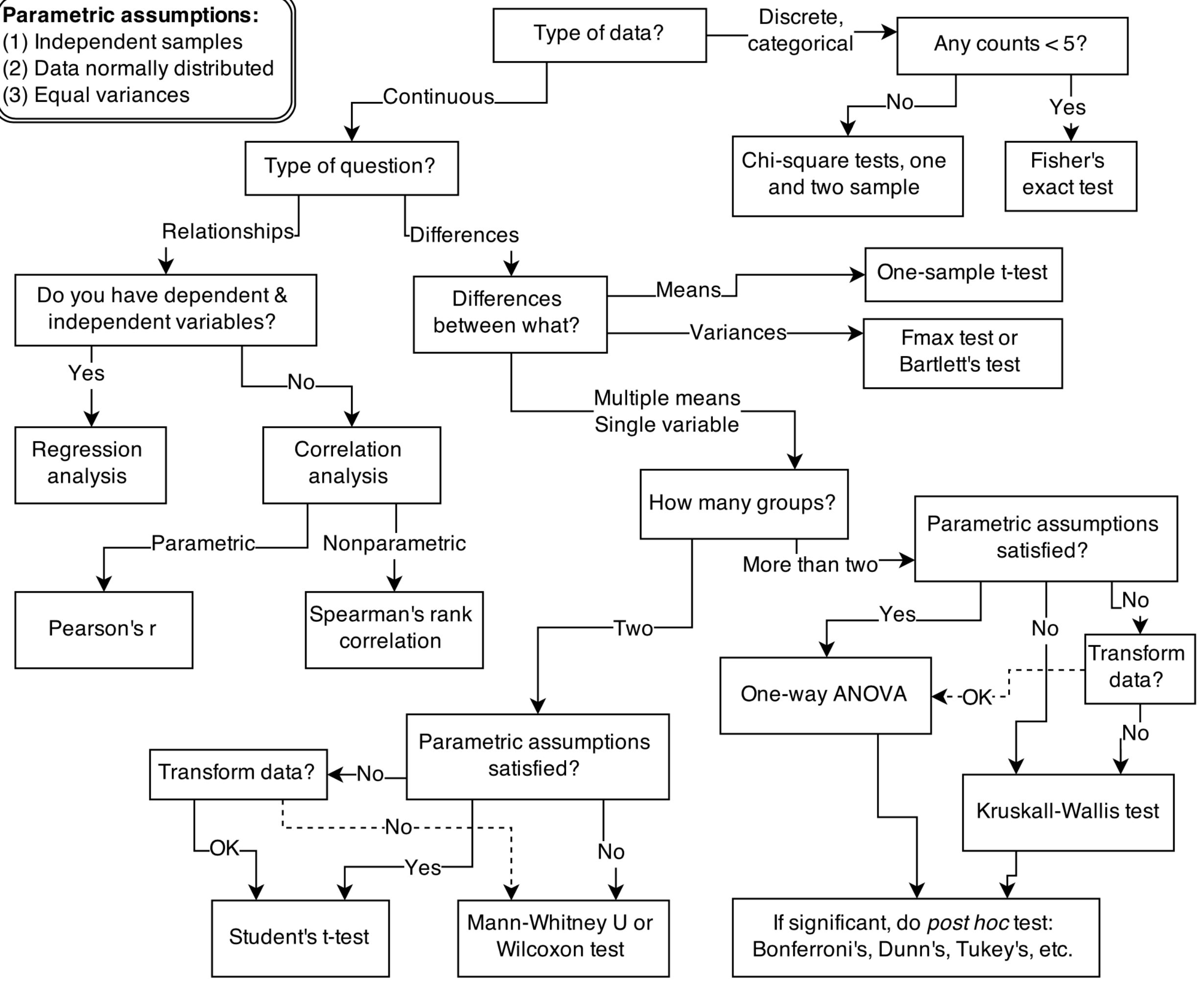
$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}, \nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{s_1^2/n_1}{n_1 - 1} + \frac{s_2^2/n_2}{n_2 - 1}}, p = F_T(t, \nu)$$

```
t.test(x, y, alternative = "two.sided", var.equal = FALSE, conf.level = 0.95)
```

zoo & decisions, comparing locations, data as $\mathbb{R} \in [0,1]$

Proportions are distributed as binomial, which tends to approximate a normal with sufficient n

Parametric assumptions:
 (1) Independent samples
 (2) Data normally distributed
 (3) Equal variances



comparing observed proportion to probability

$$H_0 : \pi = \pi_0, H_A : \pi \neq \pi_0$$

$$z = \frac{\hat{p} - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}}, p = F_{\Phi}(z)$$

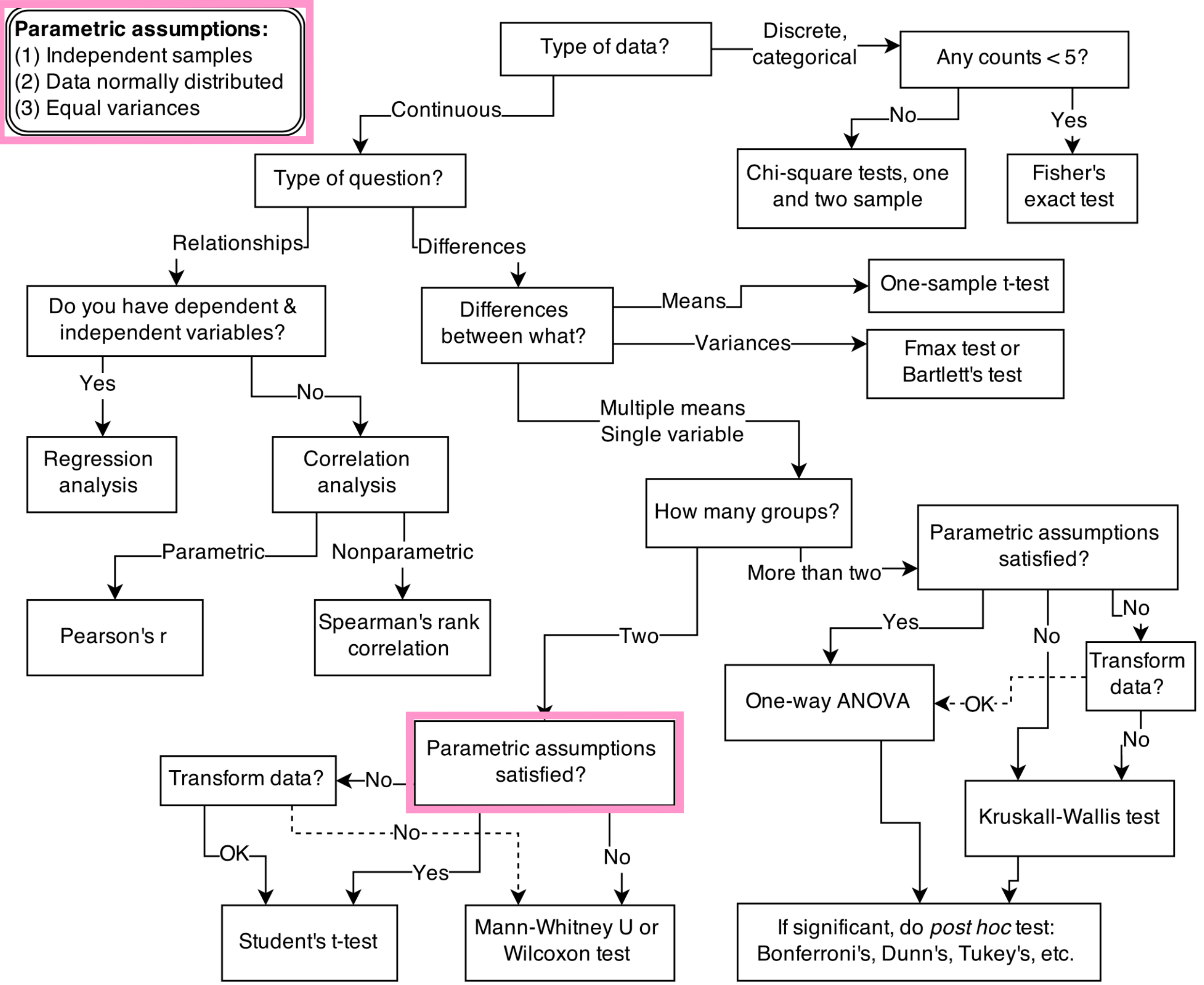
comparing two observed proportions

$$H_0 : \pi_1 - \pi_2 = \delta, H_A : \pi_1 - \pi_2 \neq \delta$$

$$z = \frac{\hat{p}_1 - \hat{p}_2 - \delta}{\sqrt{p_0(1 - p_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \text{ where } p_0 = \frac{x_1 + x_2}{n_1 + n_2}, p = F_{\Phi}(z)$$

```
prop.test(...)
```

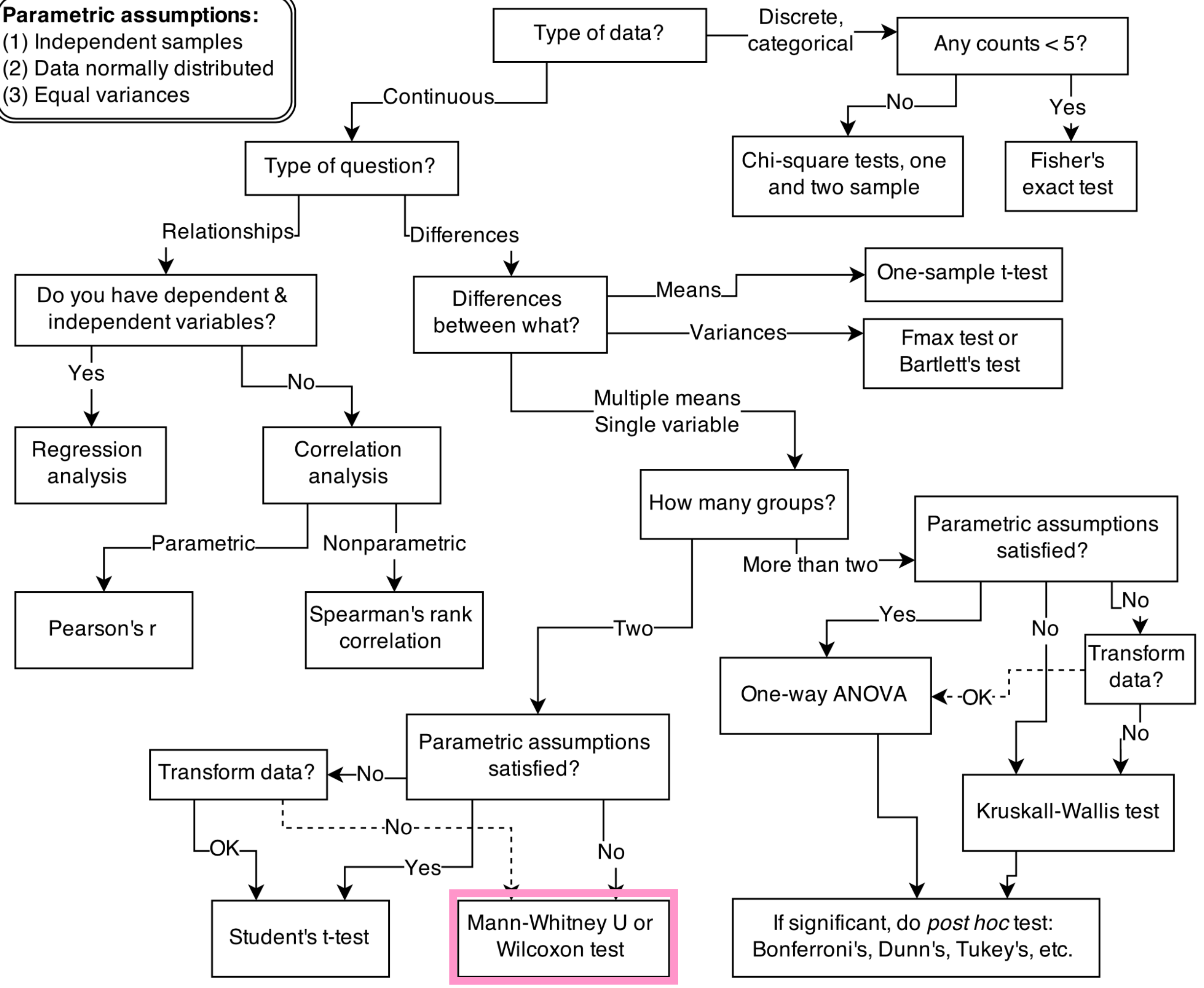
zoo & decisions, comparing locations, data as an unknown or skewed distribution



But what if we have *no reason to assume* the population is normally distributed?

zoo & decisions, comparing locations, data as an unknown or skewed distribution

Parametric assumptions:
 (1) Independent samples
 (2) Data normally distributed
 (3) Equal variances



Some tests, like *Wilcoxon–Mann–Whitney*, do not rely on parametric assumptions. *WMW* assumes independence of observations and outcomes are ordinal or continuous.

$$H_0 : P(x_i > y_j) = 1/2 \text{ (i.e., populations have same central tendency)}$$

$$H_A : P(x_i > y_j) \neq 1/2 \text{ (i.e., central tendencies not same)}$$

Simplified procedure — rank all $(n_1 + n_2)$ observations in ascending order; assign ties their average rank; sum each of the two rankings, T_a and T_b ; calculate the U statistic:

$$U_a = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - T_a, \quad U_b = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - T_b$$

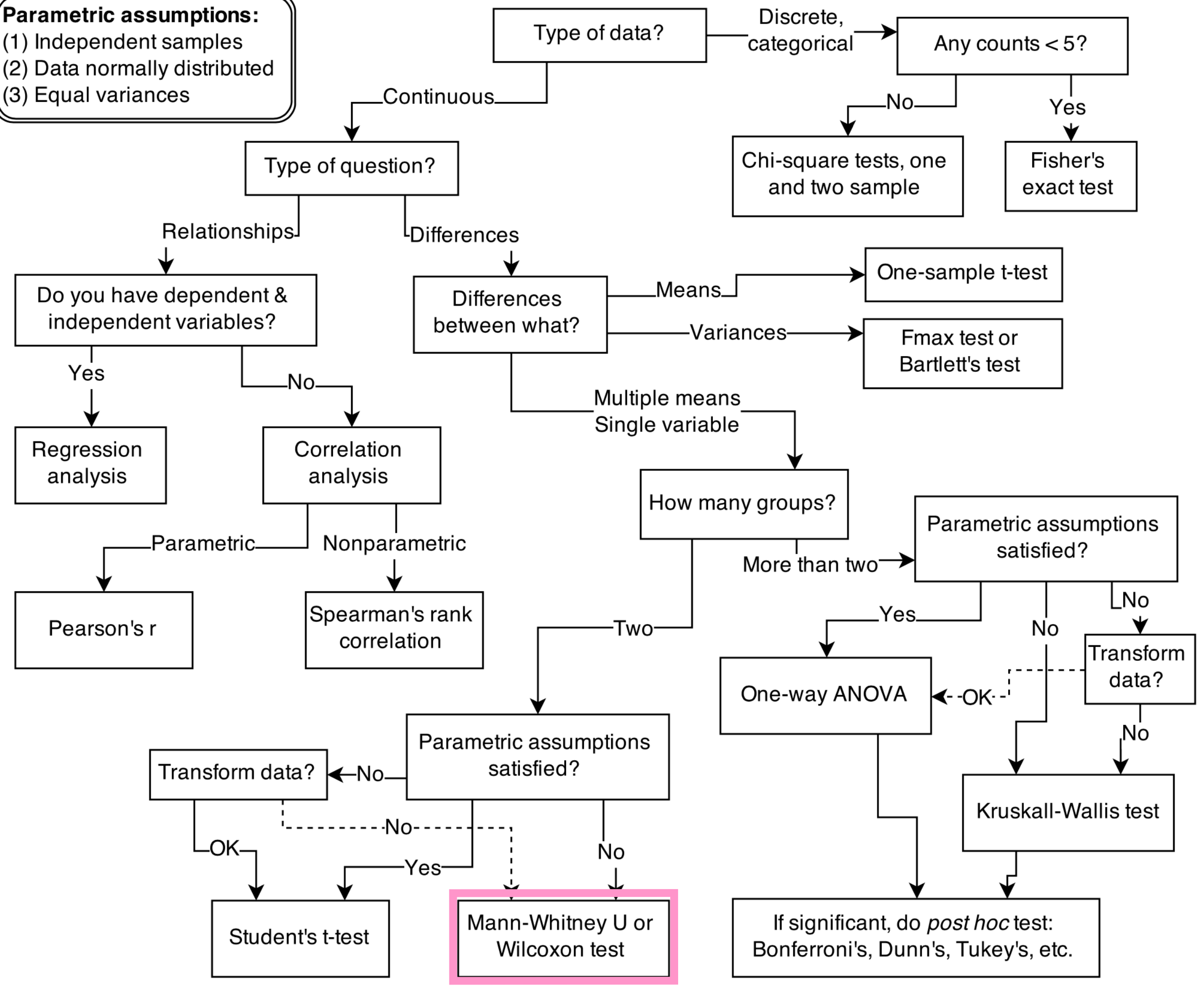
then $U = \min(U_a, U_b)$. For $n > 20$,

$$z = \frac{U - E(U)}{\sigma}, \quad E(U) = \frac{n_1 n_2}{2}, \quad \sigma^* = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}, \quad p = F_{\Phi}(z)$$

*An adjustment to σ is needed for intergroup ties.

zoo & decisions, comparing locations, data as an unknown or skewed distribution

Parametric assumptions:
 (1) Independent samples
 (2) Data normally distributed
 (3) Equal variances



Wilcoxon–Mann–Whitney, simplified code example:

```

# simulate samples from experiment, samples from different distributions
n <- 10

set.seed(1)
sample1 <- rbeta(n, 2, 2)
sample2 <- rbeta(n, 2, 5)

d <-
  data.frame(sample = rep(1:2, each = n),
             values = c(sample1, sample2)) %>%
  arrange(values) %>%
  mutate(order = seq(nrow(.))) %>%
  group_by(values) %>%
  mutate(rank = mean(order))

Ta <- filter(d, sample == 1) %>% .$rank %>% sum()
Tb <- filter(d, sample == 2) %>% .$rank %>% sum()

n1 <- with(d, sum(sample == 1))
n2 <- with(d, sum(sample == 2))

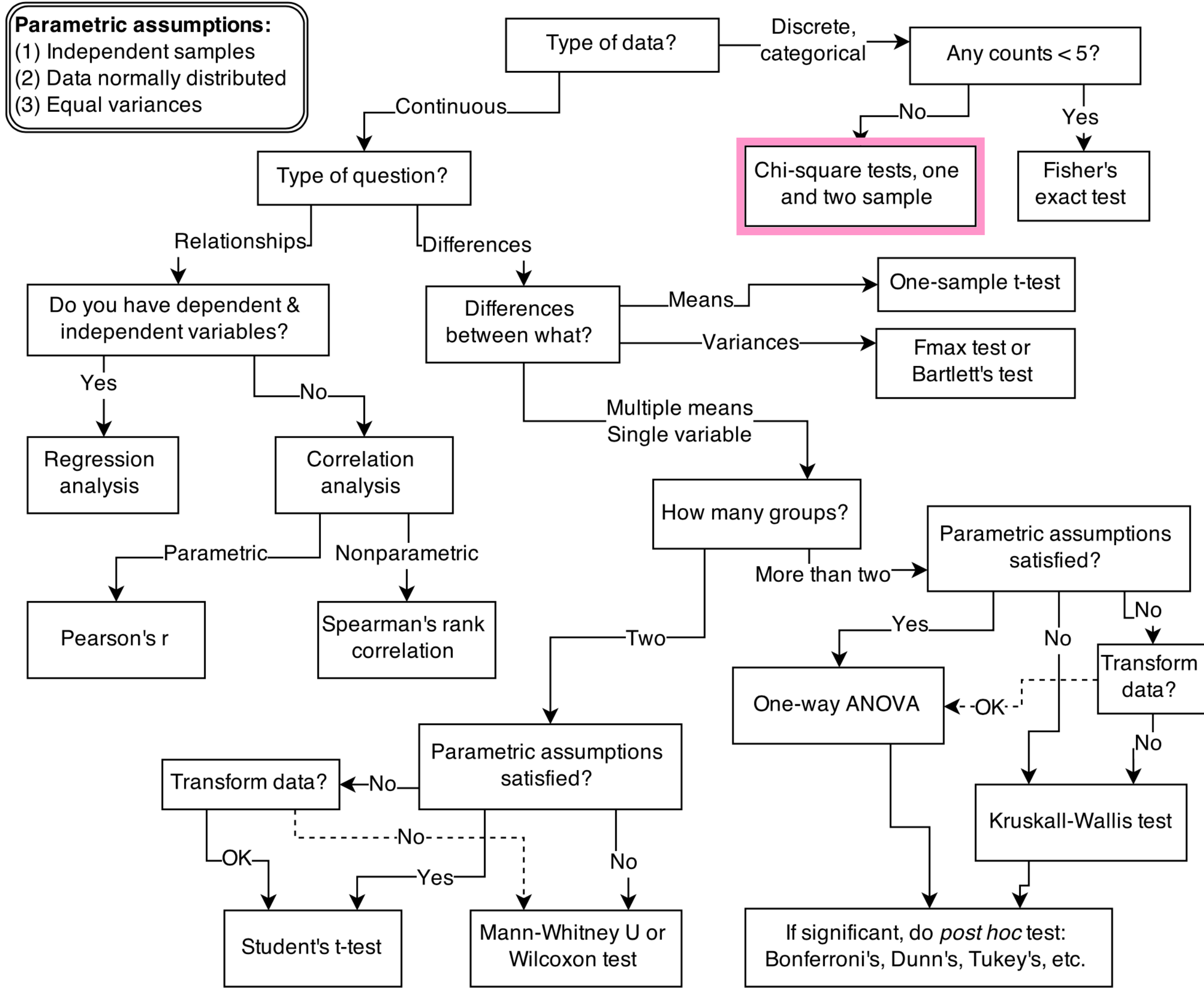
Ua <- n1 * n2 + (n1 * (n1 + 1)) / 2 - Ta
Ub <- n1 * n2 + (n2 * (n2 + 1)) / 2 - Tb
U <- min(Ua, Ub)

EU <- n1 * n2 / 2
sigma <- sqrt( n1 * n2 * (n1 + n2 + 1) / 12 )

z <- (U - EU) / sigma
p <- pnorm(z)

# compare with base R test, setting parameters to above simplifications
wilcox.test(x = sample1, y = sample2,
            correct = FALSE, exact = FALSE,
            alternative = "greater")
    
```

zoo & decisions, goodness of fit



Tests can suggest whether a whole probability distribution fits some count of categorical outcomes.

H_0 : distribution fits data

H_A : distribution doesn't fit data

$k \in \{1, \dots, K\}$ outcome categories

O_k observed counts for category k

p_k probability of category k

$E_k = n \cdot p_k$, expected counts for category k

w test statistic, variations from expected counts

$$w = \sum_{k=1}^K \frac{(O_k - E_k)^2}{E_k}, \quad \nu = K - 1, \quad p = F_{\chi^2}(w, \nu)$$

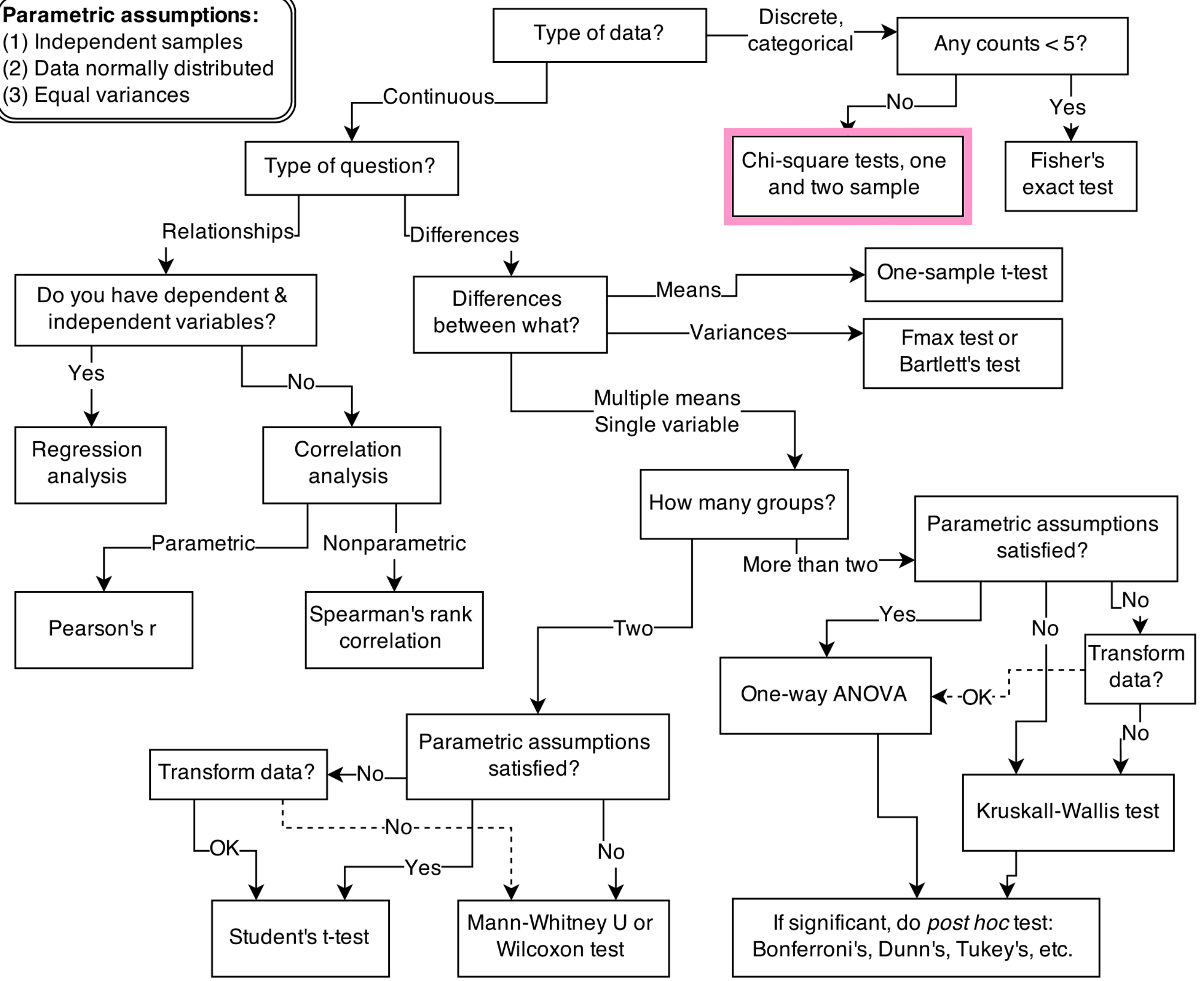
```
# Example - test equality of proportions of male and female applicants to Berkeley

library(dplyr)
data(UCBAdmissions)

UCBAdmissions %>% as.data.frame() %>%
  group_by(Gender) %>%
  summarise(O = sum(Freq)) %>%
  ungroup() %>%
  mutate(E = mean(O)) %>%
  summarise(w = sum((O - E) ^2 / E),
            nu = n_distinct(Gender) - 1) %>%
  mutate(p = pchisq(w, nu, lower.tail = FALSE))
```

zoo & decisions, independence

Parametric assumptions:
 (1) Independent samples
 (2) Data normally distributed
 (3) Equal variances



Tests can suggest whether variables are independent.

$$H_0 : P(Y|X) = P(Y) \text{ and } P(X|Y) = P(X)$$

$$H_A : P(Y|X) \neq P(Y) \text{ or } P(X|Y) \neq P(X)$$

$k \in \{1, \dots, K\}$ categories

$j \in \{1, \dots, J\}$ different levels in each category

O_{jk} observed counts for each level j in category k

p_k probability of category k

$$E_k = n \cdot p_k, \text{ expected counts for category } k$$

w test statistic, variations from expected counts

$$w = \sum_{j=1}^J \sum_{k=1}^K \frac{(O_{jk} - E_{jk})^2}{E_{jk}}, \quad \nu = (J - 1)(K - 1), \quad p = F_{\chi^2}(w, \nu)$$

Example - $H_0 : P(\text{Admit} | \text{Gender}) = P(\text{Admit})$ and $P(\text{Gender} | \text{Admit}) = P(\text{Gender})$

```

UCBAdmissions %>% as.data.frame() %>%
  mutate(Admit_pct = sum(ifelse(Admit == "Admitted", Freq, 0)) / sum(Freq)) %>%
  group_by(Gender) %>%
  mutate(E = sum(Freq) * ifelse(Admit == "Admitted", Admit_pct, 1 - Admit_pct)) %>%
  group_by(Gender, Admit) %>%
  summarise(O = sum(Freq),
            E = mean(E)) %>%
  ungroup() %>%
  summarise(w = sum((O - E)^2 / E),
            nu = (n_distinct(Admit) - 1) * (n_distinct(Gender) - 1)) %>%
  mutate(p = pchisq(w, nu, lower.tail = FALSE))
    
```

describing variation in our tests: confidence intervals

confidence intervals

$$\left[(\bar{X} - \bar{Y}) + t_{\alpha/2}\sigma, (\bar{X} - \bar{Y}) + t_{(1-\alpha)/2}\sigma \right]$$

group project work!

References

Blitzstein, Joseph K., and Jessica Hwang. *Introduction to Probability*. Second edition. Boca Raton: Taylor & Francis, 2019.

Casella, George, and Roger L. Berger. *Statistical Inference*. 2nd ed. Australia; Pacific Grove, CA: Thomson Learning, 2002.

Gelman, Andrew, Jennifer Hill, and Aki Ventari. *Regression and Other Stories*. S.l.: Cambridge University Press, 2020.

Lehmann, E L, and George Casella. *Theory of Point Estimation*. Second. Springer, 1998.

Lehmann, E. L., and Joseph P. Romano. *Testing Statistical Hypotheses*. 3rd ed. Springer Texts in Statistics. New York: Springer, 2005.

McElreath, Richard. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 2nd ed. CRC Texts in Statistical Science. Boca Raton: Taylor and Francis, CRC Press, 2020.