Non-Parametric Tests

One	Goodness of Fit	`chisq.test(x, y = NULL,p)`
	Independence Test	`chisq.test(table)` `cramer_v(table)`
	Binomial Test	`binom.test(x, n, p = 0.5)`
	Runs Test	`RunsTest(vector)`
	KS Test	`ks.test(x,dist)`
	Kendall's Tau	`cor.test(x,y,method = 'kendall')`
	Kendall's W	`kendall(df)`
Two Dep	Wilcoxon Signed Ranks	`wilcox.test(x,y,paired=T)`
	McNemar Test	`mcnemar.test(matrix)`
	Sign Test	`binom.test(condition, n, p)`
Two Indep	Mann-Whitney U	`wilcox.test(x,y,paired=F)`
	KS Z Test	`ks.test(x,y)`
	Moses Extreme Reaction	`MosesTest(x,y)`
	Wald-Wolfowitz Runs	`RunsTest(x,y)`
K Dep	Friedman Test	`kendall(df)`
	Cochran's Q Test	`cochran_qtest(data,DV~Group|Subj)`
K Indep	Kruskal–Wallis	`kruskal.test(x,g)`
	Jonckheere-Terpstra Trend Test	`JonckheereTerpstraTest(x,g)`
	Median Test	`Median.test(x.g)`

Non-parametric tests: distribution-free, less powerful
Data: skewness, outliers, tiny samples, nominal & ordinal
Chi-square: df=category-1, cell sample>5
Kendall's W: concordance-W, difference-χ^2
Wilcoxon: **less** than critical, **swap** and **smaller**

ANCOVA

Equal Variance	`levene_test(y~g)`
Independence	`anova(aov(Cov~IV,data))`
Equal Slope	`anova(aov(DV~Cov*IV,data))`
ANCOVA	`aov(DV~Cov+IV,data);` `Anova(model, type = '3')`
Effect Size	`partial_eta_squared(model)`
Post-Hoc	`summary(glht(cfit,` `linfct=mcp(time='Tukey')))`

If more grouping variable, say a &b
- equal variance: `levene_test(y~a*b)`
- independence: `anova(aov(cov~a*b))`
- equal slope: `anova(aov(y~c*a*b))`
- ANCOVA: `anova(aov(y~c+a+b+a*b))`

Covariate: continuous, not manipulated
Latin Square: less error; more power; major & minor with same level & one-time cross; no need for interaction check

Nested ANOVA

Fit Model	`aov(DV~group/subgroup,data)` `aov(DV~group+Error(group:subg),data)`

Subgroup's effect isn't care-worthy.

MANOVA

Normality	`select(DV) |> mshapiro_test()`
Var & Cov	`select(DV) |>` `box_m(group)+leveneTests(group)`
Correlation	`Corr(DV)`
MANOVA	`manova(DV~group),summary(model)`
Effect Size	`rstatix::eta_squared(model)`

Univariate ANOVA	`summary.aov(model)`

Post-Hoc	`gather(key='var',value='value',DVs) |>`
	`group_by(var) |>`
	`games_howell_test(value~group)`

Report:
- ANOVA table: F, p, eta-squared
- Post-Hoc test: MD, p

Wilk's Lambda: residuals' proportion in variance
MANOVA: less powerful, contain inflation of α

Logistic Regression

Fit Model	`lm(DV~IV+IV,data, family = binomial('logit'))`
Effect Size	`exp(cbind(coef(fit),confint(fit))`
Predictor Test	`wald.test(Sigma = vcov(fit),b = coef(fit),Terms = i:j)`
Forward Test	`anova(fit,test='Chisq')`
Null Test	`anova(fit,null,test='Chisq'),null-DV~1`
Model Fitting	`hoslem.test(x = actual, y=fitted, g)`
R Square	``pscl::pR2(fit), `r2CU`-Nagelkerke R2``
Model Comparison	`anova(fit1,fit2,test='Chisq')`
Classification Table	`confusionMatrix(round(fitted),actual)`

Rate the Model:
- Nagelkerke R^2: variation accounted for
- HL test: goodness of fit

Report:
- Analysis method + brief model intro.
- Comparison against null; R squared
- Classification table
- Wald test of coefficients
- Effect size (odds ratio)

Predicted \ Actual	1	0
1	True positive	False positive
0	False negative	True negative

$$Acc = \frac{TP+TN}{ALL}, Prec = \frac{TP}{pred\ P}, Sen = \frac{TP}{real\ P}, Spe = \frac{TN}{actual\ N}$$

MLR

Linear Correlation	`Corr(), pcor(), spcor(), vif(model)`
Fit Model	`lm(DV~IV+IV,data)`
Standardized	`lm(scale(DV)~scale(IV1),data)`
Model Comparison	`anova(fit1,fit2)`
Model's info: coefficients, confint, residuals, fitted, deviance	

Rate the Model:
- R^2_{adj}
- F & p: predicted & actual

Hypothesis Testing:
- Null and Alternative hypothesis
- Statistics (comparison if necessary)
- Conclusion (whether or not to reject)

Part correlation:
- **Unique** contribution, change in R^2 after an IV.
- **Smaller** than partial correlation.
- Read by **ROW**

Moderation

Center the Data	`C_var=var-mean(var)`
NULL Model	`lm(y~x+m)`
Moderation	`Lm(y~x*m)`
After Significant	`model_moderation <- lm(Loyalty ~ Relationship*age_con)` `emmeans(model_moderation, ~ Relationship*age_con, cov.keep = 3, at = list(age_con = c( m_Age-sd_Age, m_Age, m_Age+sd_Age), Relationship = c(m_Relationship-sd_Relationship,m_Relationship, m_Relationship+sd_Relationship)), level = 0.95)`
	`emtrends(model_moderation, pairwise ~ age_con, var = "Relationship",cov.keep = 3, at = list(age_con = c(m_Age-sd_Age, m_Age, m_Age+sd_Age)), level = 0.95)`
	`emmip(model_moderation, age_con ~ Relationship, cov.keep = 3, at = list( Relationship = c(m_Relationship-sd_Relationship,m_Relationship, m_Relationship+sd_Relationship), age_con = c(m_Age-sd_Age, m_Age, m_Age+sd_Age)), CIs = TRUE, level = 0.95, position = "jitter")`
Residual Plot	`car::residualPlots(model_moderation)`

Moderator influence the direction and strength of the relationship between IV and DV.

Report:
- Moderation: t, p, conclusion
- Main effect: t, p, conclusion
- Multicollinearity
- Model comparison for interaction term

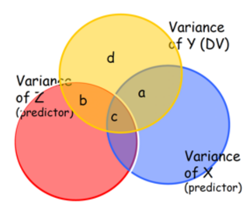
Mediation

Y on X	`Model_0=lm(Y~X)`
M on X	`Model_M=lm(M~X)`
Y on X & M	`Model_Y=lm(Y~X+M)`
Mediation Effect	`mediation::mmdiate(Model_M, Model_Y,treat=X,mediator=M, boot=T,sims=500)`

* ACME, average causal mediation effect, indirect effect.
* ADE: average direct effect.
* Total Effect: sum of direct and indirect effect.
* Proportion Mediated: percentage of indirect within total.

Report:
- Mediation; Direct effect
- Partial/complete mediator, proportion mediated

Moderation & Mediation is linear relationship essentially.



- Partial: Holding Z from BOTH X and Y; Good for knowing the variance uncounted by other control variables.
- Semi-Partial: Holding Z from X only; Good for knowing the unique contribution of X.

Distance

Correlation	`cor()<v>`
Euclidean	`data %>% scale() %>% dist()<c>`
Cosine	`data %>% lsa::cosine()<v>`

PCA

Bartlett Sphericity	`bfi_cor=cor(bfi)` `cortest.bartlett(bfi_cor,nrow(bfi_cor))`
KMO MSA	`KMO(bfi_cor)`
No. of Factors	`scree(bfi),fa.parallel(bfi)`
Factor Analysis	`fa(r,nfactors,rotate,fm)`
Factor Diagram	`fa.diagram(model)`

KMO: variance explained by common factors
Loading: correlation coefficient; Weight: slope coefficient.
Communality: variance explained by selected factors.
Item complexity: each item loads on how many factors.
SS loadings: eigenvalues, sum of squared loadings.
Eigenvalue and communality change after rotation. Total variance explained unchanged.
Varimax maximizes variance accounted for by single factor; enhance interpretability. Variable links with min No. of factors. New factors are uncorrelated (Orthogonal).
Sample greater than 5 times items. Only continuous data.

Reliability Analysis

Reliability	`jmv::reliability(data,vars,alphaScale,` `corPlot,revItems,alphaItems,itemRestCor)`

Nonadditivity Test	`athletes %>% Reverse_Coding %>%`
	`pivot_longer(cols =everything(),n`
	`ames_prefix = 'x',names_to = 'item',`
	`values_to = 'score') %>%`
	`  mutate(subj =`
	`rep(1:nrow(athletes),each =`
	`ncol(athletes)))`

	`nonadd = lm(score`
	`~subj+as.factor(item),athletes_long)`
	`MSerror`
	`=deviance(nonadd)/df.residual(nonadd)`
	`res =  with(athletes_long,`
	`nonadditivity(score,subj,item,`
	`df.residual(nonadd), MSerror))`

ICC	`icc=ICC(data,check.keys=T)` `icc$results,icc$stats`

More items, higher Cronbach's alpha.
Higher ICC, higher consistency within group (subj).
Qualitative: consistency; Quantitative: absolute aggrement
ICC1: random, missing data.

ICC Terms	Only the subject (item) is the random factor. Not every rater measures every subject. Ratings are randomly drawn from a population	
One-way Random Effects	ICC(1) AND the effects of rater and subject (item) are mixed together	
TYPE:	Consistency	Absolute Agreement
Two-way Random Effects Sample	ICC(2) Consistency does not mean agreement	difference not matter→ consistency; absolute agreement ICC(2)
Two-way Mixed Model: Raters Fixed, Subjects Random Population	ICC(3) Not interested in generalization to raters	ICC(3) ICC3 greater than ICC2

Single: accuracy of each rater; Mean: overall consistency
R code: ICC2-abs agreement; ICC3-consistency

Survival Analysis

Survobj	`Surv(time,event)`
Survival Model	`survfit(Survobj~1,data)` `survfit(Survobj~group,data)`
	`coxph(Survobj~as.factor(g),dt,model=T)` `survdiff(Survobj~group,data)`
Survival Plots	`ggsurvplot(model,data,fun='pct')`

Median is used, since not everyone will have an event.
Hazard function: (ds/s)/dt

Cox regression: $h(t)=h0(t)*exp(B0+B1X)$; B1:hazard ratio

Clustering

Hierarchical	`distMat=data %M% scale() %M% dist()`
	`hc=hclust(distMat)`
	`plot(hc,hang=-1);fviz_dend(hc,k,rect=T)`
	`memlab=cutree(hc,k);table(memlab)`
	`aggregate(data,list(memlab),mean)`

Non-Hierarchical	`kmeans(data,centers)` `fviz_cluster(kmeanmodel,data)`

No. of Clusters	`fviz_nbclust(data,kmneas,method='wss')`
	`fviz_nbclust(data,kmneas,method='silht')`
	`nclust=n_clusters(data.frame(data))`
	`print(nclust)`

FA: Correlation; Cluster: heterogeneity (No prior)
No. of clusters depends on measures of distance.
Non- and hierarchical clustering rates reliability.
Dendrogram: vertical heights imply dissimilarity

Attention:

```
matrix(data, nrow, byrow, dimnames = list(row,column))
```

Range for continuous variables
Factor: Grouping variables; set as `as.factor(var)` in `lm`
Library: psych, survival, jmv, emmeans, mediation, survminer, ggfortify, parameters, factoextra, cluster, tidyverse, rstatix, DescTools, car, effectsize, irr, bruceR, aod, ggplot2, ResourceSelection, heplots, pscl , ppcor, HH, agricolae, multicomp, caret