Logistic Regression & Survival Analysis
Statistical Data Analysis using SPSS

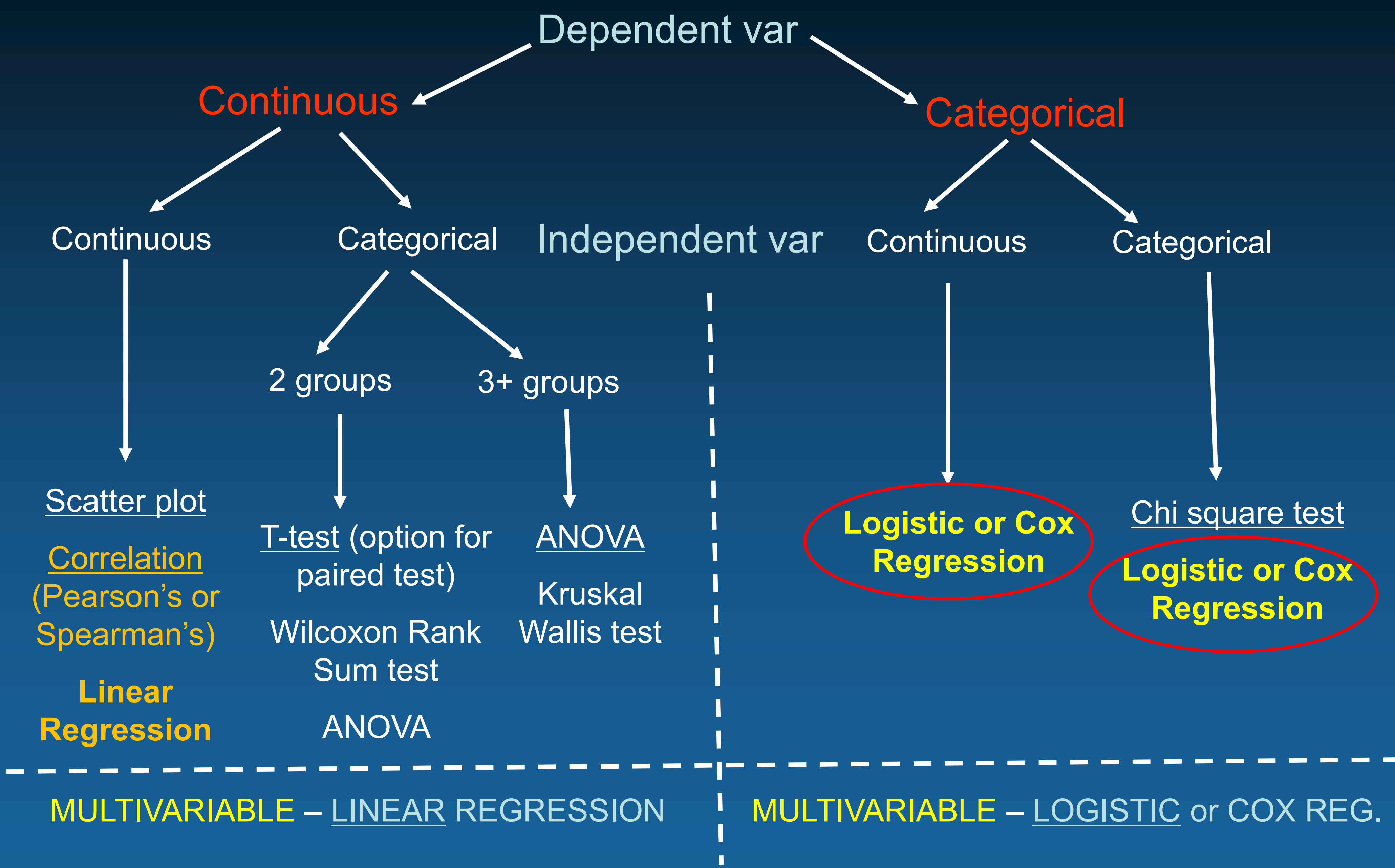
Ilir Agalliu MD, Sc.D
Associate Professor
Dept. of Epidemiology & Population Health

Pediatrics Fellows 3rd Year
9/20/2023

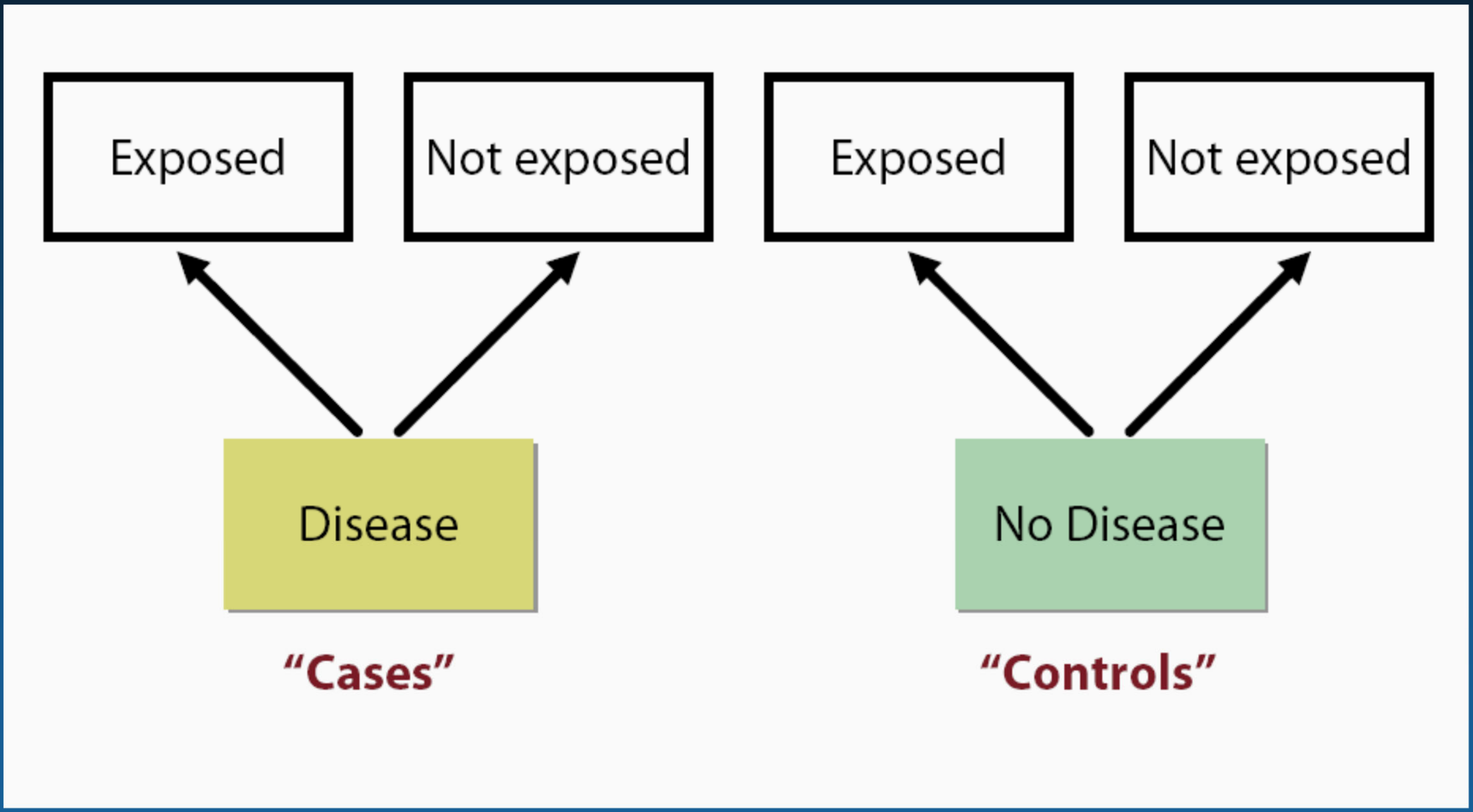
Outline

- Logistic Regression
  - When to use logistic regression?
  - The Model and Logistic Function
  - Interpretation for Indicator Variables
  - Evaluation of the Model and Examples
- Survival Analysis
  - When to use survival analysis?
  - Kaplan Meier Function & Cox PH Model
  - Examples
- Statistical Data Analyses using SPSS

Decision: Bivariable Analysis



Case-Control Study Design



Analysis of Case-Control Study

First, select

Then, measure past exposure

	Cases (with disease)	Controls (without disease)
Were exposed	a	b
Were not exposed	c	d
Totals	a + c	b + d
Proportion exposed	$\frac{a}{a+c}$	$\frac{b}{b+d}$

Measure of Association
Odds Ratio (OR)

	Cases	Controls
Exposed	A	B
Not Exposed	C	D

Odds of Exposure in Cases = A / C
Odds of Exposure in Controls = B / D

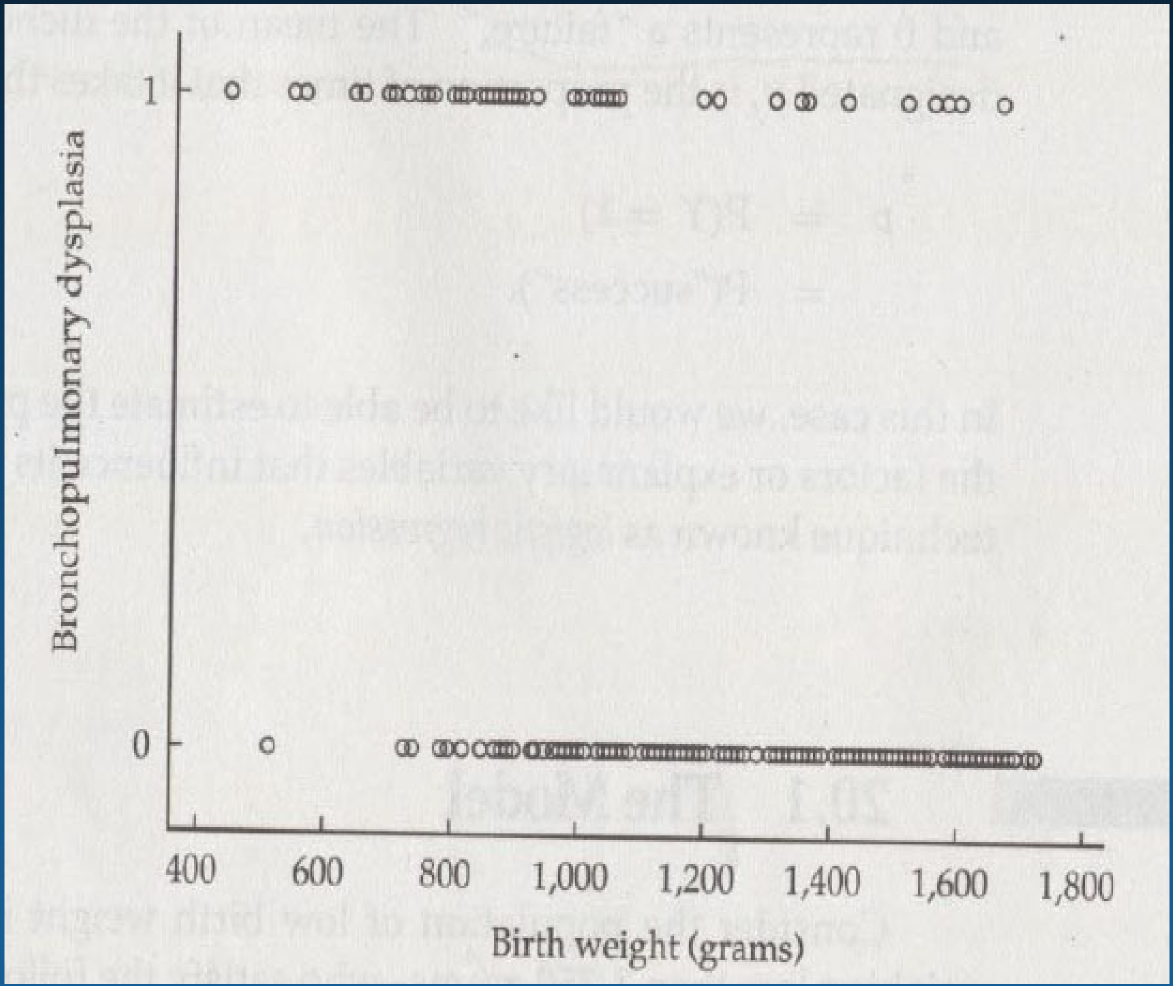
OR = A / C ÷ B / D = AD / BC

Logistic Regression

Outcome is dichotomous:
1=yes; 0=no

What is the association between broncho-pulmonar displasia (BPD) and baby's birth weight (cont)?

What do you observe in fig?
Can we fit a linear regression model here?



BPD and baby's birth weight

Birth Weight (grams)	Sample Size	Number with BPD	$\tilde{p}$
0–950	68	49	0.721
951–1,350	80	18	0.225
1,351–1,750	75	9	0.120
	223	76	0.341

One way to express birth-weight is to create categories…

What is the association between broncho-pulmonar displasia (BPD) and baby's birth weight (category)?

If you compare proportions (ρ) of lowest to highest birth-weight categories: RR = 0.721 / 0.12 = 6.01

How do you interpret this?

The Logistic Function
BPD and baby's birth weight

One might attempt to fit the model like linear regression:

$$p = \alpha + \beta_1 x_1$$

Issue: P- is probability of success and thus can take values 0 to 1

From the above equation *p* can take any value (not appropriate)

To accommodate the constrain that *p* should be 0-1 we fit equation in the form:

$$p = \frac{e^{\alpha + \beta_1 x_1}}{1 + e^{\alpha + \beta_1 x_1}}.$$

This is called the logistic function

This can also be expressed algebraically in the form:

$$\ln\left[\frac{p}{1-p}\right] = \alpha + \beta_1 x_1.$$

- where p/(1-p) the odds of success

This is called logistic regression model

Logistic Regression
BPD and baby's birth weight

Equation: $$\ln\left[\frac{\hat{p}}{1-\hat{p}}\right] = a + b_1 x_1.$$

a: intercept

b: estimate of slope

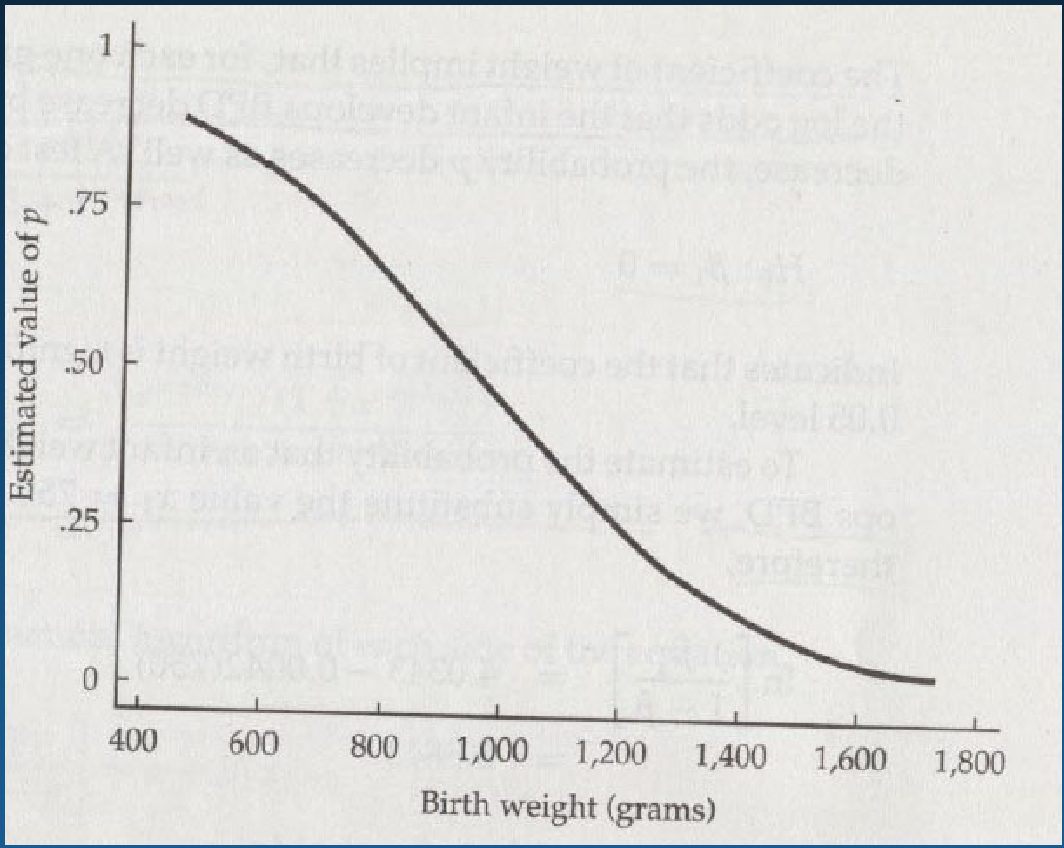
To fit logistic model, we apply maximum likelihood method

BPD and baby's birth weight

$$\ln\left[\frac{\hat{p}}{1-\hat{p}}\right] = 4.0343 - 0.0042x_1.$$

How do you interpret this?

What is the null hypothesis here?



Logistic Regression

BPD and baby's birth weight

Equation:

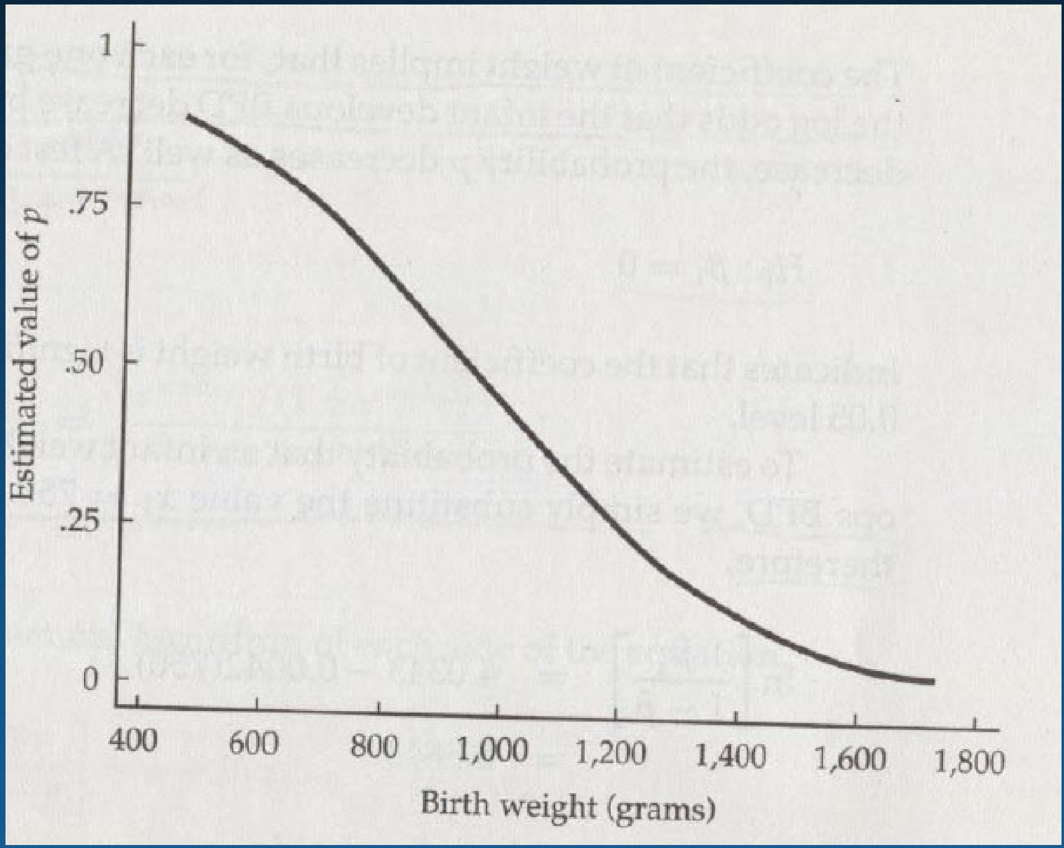
$$\ln\left[\frac{\hat{p}}{1-\hat{p}}\right] = 4.0343 - 0.0042x_1.$$

What is probability of BPD for a baby weighting 750 g at birth?

$$\begin{aligned}\ln\left[\frac{\hat{p}}{1-\hat{p}}\right] &= 4.0343 - 0.0042(750)\\ &= 0.8843.\end{aligned}$$

$$\begin{aligned}\frac{\hat{p}}{1-\hat{p}} &= e^{0.8843}\\ &= 2.4213.\end{aligned}$$

$$\begin{aligned}\hat{p} &= \frac{2.4213}{1+2.4213}\\ &= \frac{2.4213}{3.4213}\\ &= 0.708.\end{aligned}$$



What is the odd of BPD for a baby weighting 750 g at birth?

Multiple Logistic Regression

What if we want to assess simultaneously the effect of two or more predictor variables on a dichotomous outcome?

Consider the following research question

- What is the association between BPD and baby's weight and gestational age (week)?

- We can extend logistic regression to accommodate two or more independent variables:

$$\ln\left[\frac{\hat{p}}{1-\hat{p}}\right] = a + b_1x_1 + b_2x_2$$

- Same assumptions apply also for multiple logistic model
- Use the maximum likelihood method to fit the model

Logistic Regression and Indicator Variable
BPD and mother's toxemia during pregnancy

The logistic regression model can be generalized to include explanatory variables that are dichotomous (1:yes, 0:no)

$$\ln\left[\frac{\hat{p}}{1-\hat{p}}\right] = -0.5718 - 0.7719x_3.$$

In this case the coeff β=*-0.7719* indicates the relative odds of developing BPD for children whose mothers had toxemia vs. those who did not have:

OR = $e^{-0.7719}$ = 0.46

	Toxemia		
BPD	Yes	No	Total
Yes	6	70	76
No	23	124	147
Total	29	194	223

Consider the above 2 x 2 table

What is association between BPD and mother's toxemia?

OR = (6*124) / (23*70) = 0.46

An Example: Risk of Pediatric Crohn's Disease According to Antibiotic Exposure (Virta et al AJE 2012)

Outcome, Gender, and Antibiotic Use	No. of Cases	No. of Controls	Crude OR	95% CI	Adjusted[a] OR	95% CI
Crohn's disease						
None	10	75	1	Referent	1	Referent
Overall	223	857	2.18	1.03, 4.61	2.06	0.97, 4.36
Male						
None	1	46	1	Referent	1	Referent
Overall	146	542	12.67	1.73, 92.82	11.86	1.61, 87.37
Female						
None	9	29	1	Referent	1	Referent
Overall	77	315	0.74	0.31, 1.78	0.73	0.30, 1.75
No. of antibiotic purchases						
0	10	75	1	Referent	1	Referent
1–3	40	204	1.62	0.73, 3.58	1.61	0.72, 3.56
4–6	37	196	1.71	0.76, 3.86	1.68	0.74, 3.79
7–10	63	171	3.48	1.57, 7.34	3.19	1.43, 7.13
11–16	46	154	2.93	1.28, 6.68	2.70	1.18, 6.19
≥17	37	132	2.81	1.21, 6.54	2.40	1.02, 5.64
P-trend			0.001		0.009	

Logistic Regression

SPSS: Analyze -> Regression-> Binary Logistic

Is there a relationship between baby birth weight category and maternal hypertension during pregnancy, after adjusting for age and smoking during pregnancy?

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	History of hypertension(1)	1.234	.621	3.949	1	.047	3.435	1.017	11.600
	Mothers age	-.050	.032	2.409	1	.121	.951	.892	1.013
	Smoking during pregnancy(1)	.701	.326	4.627	1	.031	2.016	1.064	3.817
	Constant	-.017	.769	.000	1	.982	.983		

a. Variable(s) entered on step 1: History of hypertension, Mothers age, Smoking during pregnancy.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.790	.157	25.327	1	.000	.454

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	223.281[a]	.058	.082

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Hosmer-Lemeshow GOF Test

- As in linear regression, goodness of fit in logistic regression attempts to get at how well a model fits the data
- It is usually applied after a "final model" has been selected

- The Hosmer–Lemeshow goodness of fit (GOF) test is commonly used to assess model fit
  - The test assesses whether or not the observed event rates match the expected event rates in subgroups of the model population
- The test specifically identifies subgroups as deciles of fitted risk values
- The null hypothesis is that the model is fit
  - If the $p<0.05$ then null hypothesis is rejected => model is not fit
- Models for which expected and observed event rates in subgroups are similar are called well calibrated

Logistic Regression (Diagnostics)

Analyze -> Regression-> Binary Logistic

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8.697	8	.368

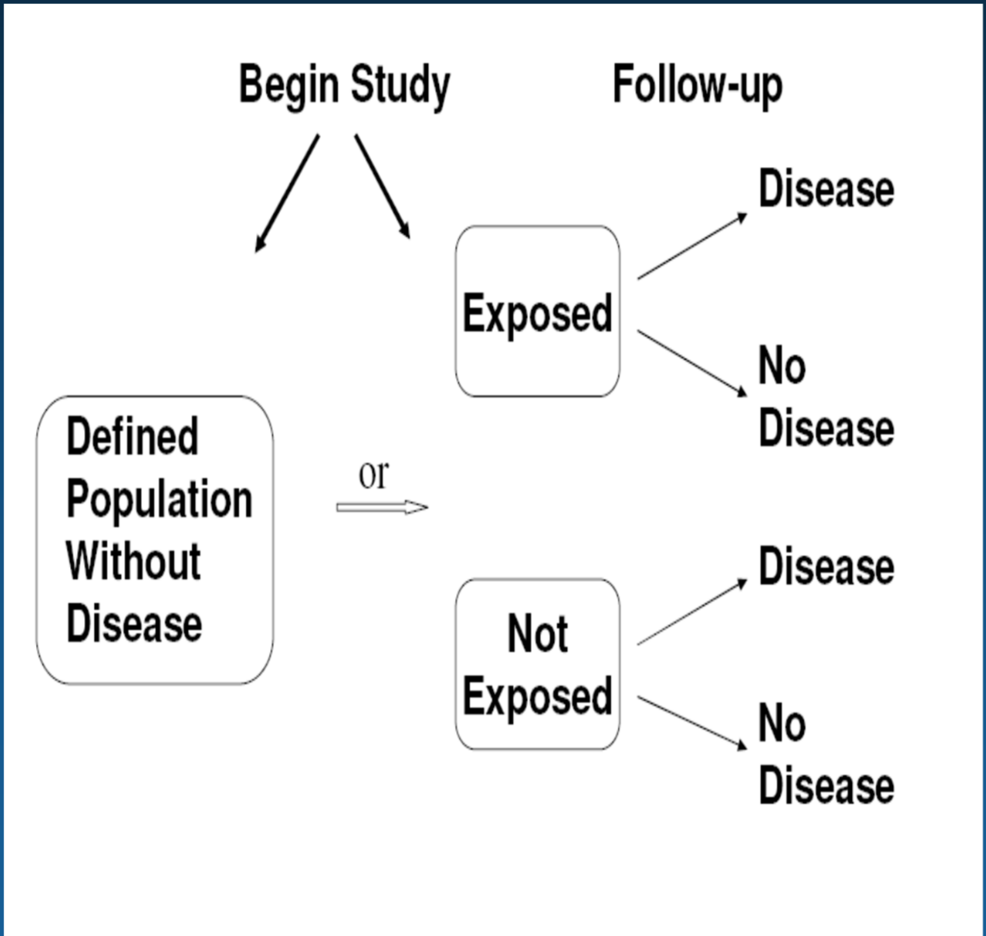
Contingency Table for Hosmer and Lemeshow Test

		Birth weight group = >2500 gm		Birth weight group = < 2500 gm		
		Observed	Expected	Observed	Expected	Total
Step 1	1	20	17.546	1	3.454	21
	2	12	15.033	7	3.967	19
	3	14	13.840	4	4.160	18
	4	18	17.859	6	6.141	24
	5	13	12.239	4	4.761	17
	6	14	13.211	5	5.789	19
	7	10	11.774	8	6.226	18
	8	10	11.914	10	8.086	20
	9	14	11.077	6	8.923	20
	10	5	5.506	8	7.494	13

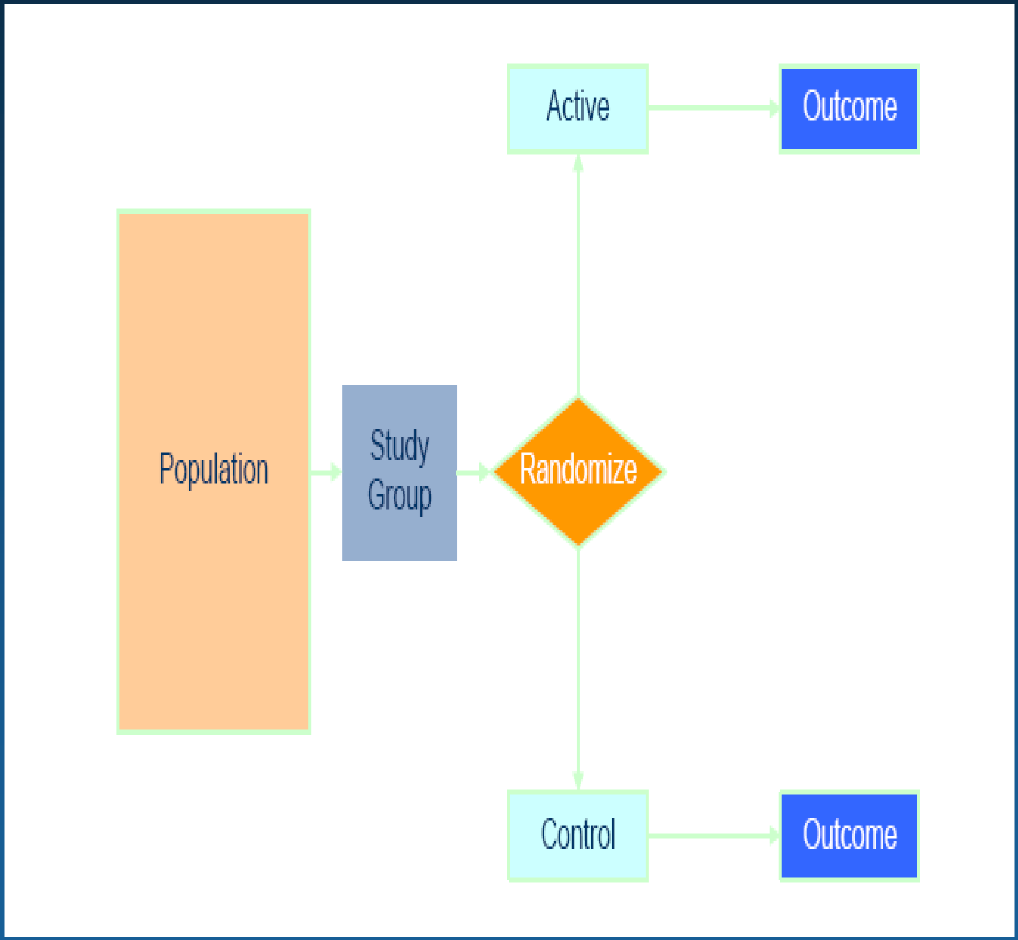
Analysis of Cohort Studies

Design of a Cohort Study

Observational



Randomized Clinical Trial



Accrual of Person-Time: Open Cohort

	Jan 1980	Jan 1989	Jan 1999	
Subject 1	--------------------D			10 Person-Years (PY)
Subject 2		E-------------------D		10 PY
Subject 3	--			20 PY
Subject 4	E------------------------------			15 PY
Subject 5	E-----------------------------X			15 PY
				70 PY (Total)

D = Diabetes, E-Entry into cohort, X- Lost to follow-up

Incident Rate of Diabetes = 2 / 70 PY

When to use Survival Analysis?

- Used to analyze data in which the time until the event is of interest
- Response is often referred to as a failure time / survival
- Examples
  - Time until tumor recurrence
  - Time until cardiovascular incidence after some treatment / intervention
  - Time until development of AIDS for HIV+ patients

Things to Consider for Survival Analysis

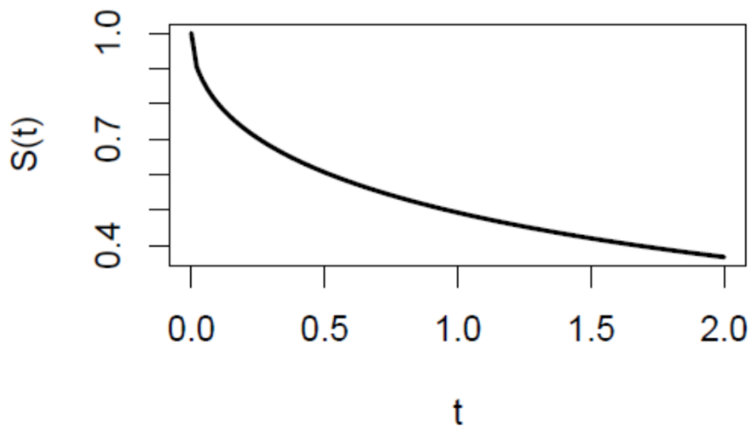
- Each subject has a beginning and an end anywhere along the timeline of a complete study
- In many clinical trials, subjects may enter or begin the study and reach end-point at vastly differing points
- Each subject is characterized by
  1. Survival time (continuous)
  2. Status at the end of the survival time (event occurrence or censored, or death)
  3. The study group they are in (e.g. placebo vs. intervention)
- Censoring
  - People who are lost to follow-up / withdraw / end of study

Survival Analysis - Terminology

- T denotes the response variable, $T \geq 0$.

- The survival function is

$$S(t) = Pr(T > t) = 1 - F(t).$$

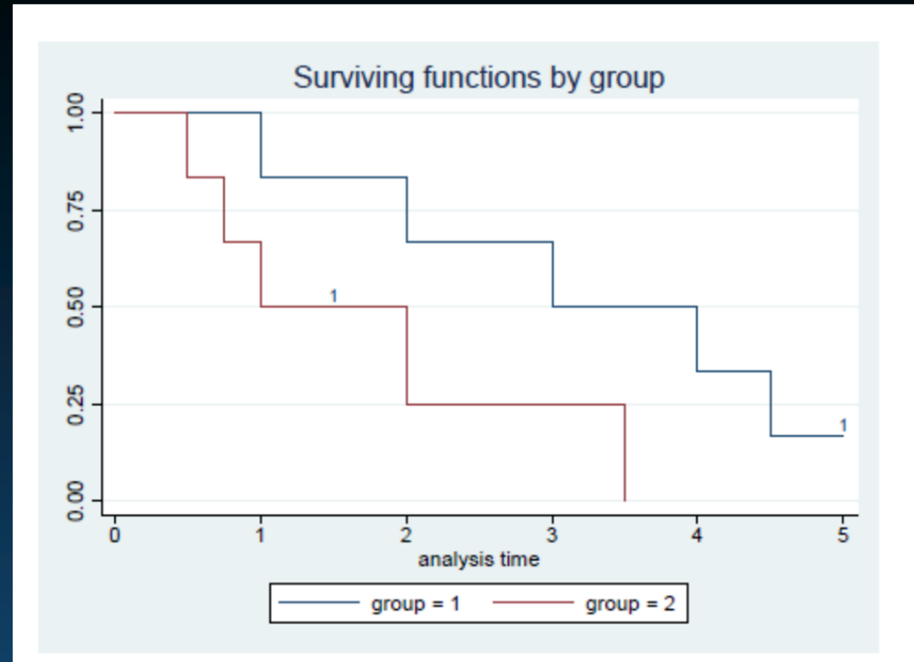


Kaplan-Meier Product-Limit Estimator

- The KM curve is a step-wise estimator, not a smooth function
- KM useful tool to visualize the difference between two survival curves
- Lengths of horizontal lines represent the survival duration for that interval
- Interval is terminated by the occurrence of the event of interest
- Vertical distances between horizontal lines illustrate the change in the cumulative probability

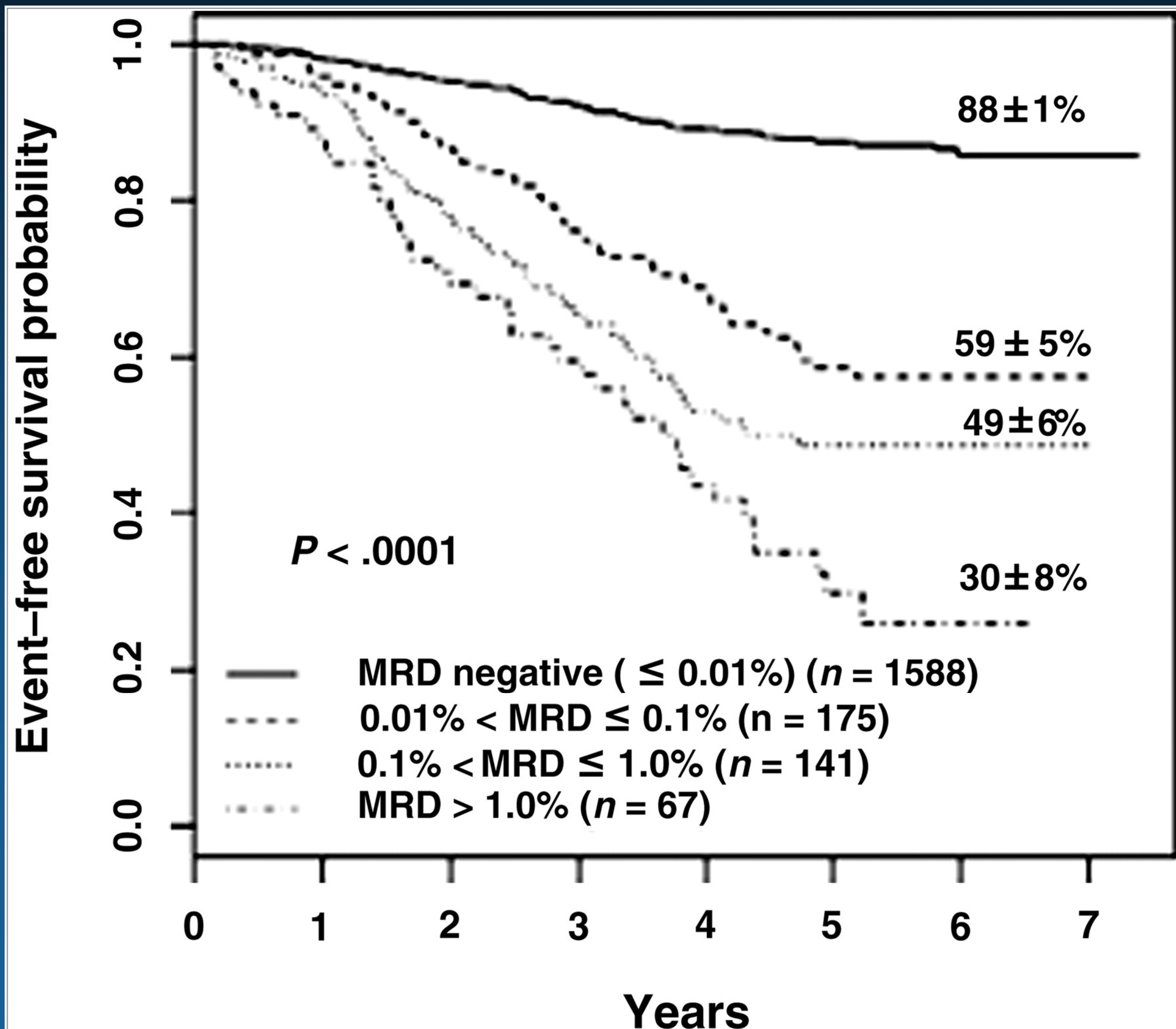
Example of KM Analysis

Comparison Log-Rank Test



Subject	Group	Survival time in the interval	# *surviving* at risk	Event	# *surviving* after event	Cumulative survival rate
1	1	1	6	1	5	$1\times\frac{5}{6}$
2	1	2	5	1	4	$1\times\frac{5}{6}\times\frac{4}{5}$
3	1	3	4	1	3	$1\times\frac{5}{6}\times\frac{4}{5}\times\frac{3}{4}$
4	1	4	3	1	2	$1\times\frac{5}{6}\times\frac{4}{5}\times\frac{3}{4}\times\frac{2}{3}$
5	1	4.5	2	1	1	$1\times\frac{5}{6}\times\frac{4}{5}\times\frac{3}{4}\times\frac{2}{3}\times\frac{1}{2}$
6	1	5		0		
7	2	0.5	6	1	5	$1\times\frac{5}{6}$
8	2	0.75	5	1	4	$1\times\frac{5}{6}\times\frac{4}{5}$
9	2	1	4	1	3	$1\times\frac{5}{6}\times\frac{4}{5}\times\frac{3}{4}$
10	2	1.5		0		
11	2	2	2	1	1	$1\times\frac{5}{6}\times\frac{4}{5}\times\frac{3}{4}\times\frac{1}{2}$
12	2	3.5	1	1	0	$1\times\frac{5}{6}\times\frac{4}{5}\times\frac{3}{4}\times\frac{1}{2}$

Example: Event-free survival (EFS) of patients receiving therapy for acute lymphoblastic leukemia with bone marrow results at the end of induction therapy (day 29) to test for minimal residual disease (MRD)



Devidas et al. Children's Oncology Group study. *Blood*. 2008;111(12):5477–5485

Cox Proportional Hazard Model

The Cox model leaves the baseline hazard function $\beta_0(t) = \log h_0(t)$ unspecified

$$\log h_i(t) = \beta_0(t) + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

The model is semiparametric, because while the baseline hazard can take any form, the covariates enter the model linearly.

- The baseline hazard does not depend on covariates, but only on time
- The covariates are time-constant
- Proportional hazard assumption follows

Prenatal Tetanus, Diphtheria, Acellular Pertussis Vaccination and Autism Spectrum Disorder

Tracy A. Becerra-Culqui, PhD, MPH, OT/L, Darios Getahun, MD, PhD, MPH, Vicki Chiu, MS, Lina S. Sy, MPH, Hung Fu Tseng, PhD, MPH

BACKGROUND: Increasing vaccination of pregnant women makes it important to assess safety events potentially linked to prenatal vaccination. This study investigates the association between prenatal tetanus, diphtheria, acellular pertussis (Tdap) vaccination and autism spectrum disorder (ASD) risk in offspring.

METHODS: This is a retrospective cohort study of mother-child pairs with deliveries January 1, 2011 to December 31, 2014 at Kaiser Permanente Southern California hospitals. Maternal Tdap vaccination from pregnancy start to delivery date was obtained from electronic medical records. A diagnosis of ASD was obtained by using *International Classification of Diseases, Ninth* and *Tenth Revision* codes. Children were managed from birth to first ASD diagnosis, end of membership, or end of follow-up (June 30, 2017). Cox proportional hazards models estimated the unadjusted and adjusted hazard ratios (HRs) for the association between maternal Tdap vaccination and ASD, with inverse probability of treatment weighting to adjust for confounding.

TABLE 2 Follow-up and ASD Diagnosis in Children Born Between 2011 and 2014 to Women Who Were Unvaccinated and Vaccinated With Tdap During Pregnancy

	Unvaccinated *n* = 42 916	Vaccinated *n* = 39 077	*P*[a]
Follow-up characteristics			
Total follow-up time (1000 person y)	190.74	150.56	—
Length of follow-up, y			
Mean (SD)	4.44 (1.18)	3.85 (1.29)	<.0001
Median	4.60	3.50	
Q1, Q3	3.7, 5.3	2.9, 4.9	
Range	(1.2–6.5)	(1.2–6.5)	
Reasons for ending follow-up			
Termination of KPSC membership, *n* (%)	6508 (15.2)	5242 (13.4)	<.0001
End of study (June 30, 2017), *n* (%)	35 636 (83.0)	33 266 (85.1)	<.0001
ASD diagnosis, *n* (%)	772 (1.8)	569 (1.5)	.0008
ASD diagnosis prevalence by birth y, *n* (%)			
2011	218 of 11 202 (1.9)	143 of 8063 (1.8)	.3836
2012	282 of 15 146 (1.9)	80 of 5407 (1.5)	.0666
2013	206 of 12 017 (1.7)	145 of 8725 (1.7)	.7729
2014	66 of 4551 (1.5)	201 of 16 882 (1.2)	.1611
ASD diagnosis age, *n* (%), y	772 (100)	569 (100)	<.0001
1	97 (12.6)	116 (20.4)	
2	337 (43.7)	251 (44.1)	
3 or 4	314 (40.7)	178 (31.3)	
5 or 6	24 (3.1)	24 (4.2)	

TABLE 3 Rates and Associations Between Tdap Vaccination During Pregnancy and ASD Among Children Born Between 2011 and 2014

	ASD Incidence Rate per 1000 Person y		HR (95% CI)	
	Unvaccinated	Vaccinated	Unadjusted	IPTW-Adjusted[a]
Overall	4.05	3.78	0.98 (0.88–1.09)	0.85 (0.77–0.95)
Birth y				
2011	3.57	3.22	0.91 (0.74–1.12)	0.86 (0.70–1.07)
2012	4.02	3.18	0.80 (0.62–1.02)	0.80 (0.63–1.03)
2013	4.48	4.46	1.00 (0.81–1.23)	0.99 (0.80–1.23)
2014	4.87	4.14	0.89 (0.68–1.18)	0.85 (0.65–1.12)
Nulliparous	4.88	4.56	0.99 (0.85–1.15)	0.88 (0.75–1.02)

[a] Adjustments were made for child's birth y, gestational age at birth (<37 or ≥37 wk); maternal age, race and/or ethnicity, and education; Medicaid insurance, medical center of delivery, parity, start of prenatal care, and influenza vaccination during pregnancy.

Take Home Messages

- Logistic regression is used to analyze categorical outcomes (e.g. case-control studies)
  - Calculates directly the probabilities of events for a set of predictor variables
  - If independent variable is a dichotomous; you can calculate directly OR
  - Check Hosmer-Lemeshow GOF test
- Survival analysis is used to model time to event data
  - Caution should be made about censoring issues
  - KM useful tool to visualize the difference between two (or more) survival curves
  - Cox PH model is used to adjust for confounding

Statistical Data Analysis using SPSS in Clinical Research

Research Questions

Using data birthwt.sav please address the following:

1. *Is there a correlation between mother's age and baby's weight at birth?*
2. *Is there a statistically significant difference in baby's birth weight (as continuous) by maternal smoking during pregnancy?*
3. *Is there a statistically significant difference in baby's birth weight (as continuous) by mother's race?*
4. *Is there a significant difference in proportions of baby birth weight groups (categorical) by maternal hypertension during pregnancy?*
5. *Is there a relationship between baby birth weight and hypertension during pregnancy, after adjusting for maternal age and smoking status?*

Data View

birthwt.sav

	ID	BW_GRP	MOTH_AGE	MOTH_WT	RACE	SMOKE
1	85.00	0	19.0	182.00	2	0
2	86.00	0	33.0	155.00	3	0
3	87.00	0	20.0	105.00	1	1
4	88.00	0	21.0	108.00	1	1
5	89.00	0	18.0	107.00	1	1
6	91.00	0	21.0	124.00	3	0
7	92.00	0	22.0	118.00	1	0
8	93.00	0	17.0	103.00	3	0
9	94.00	0	29.0	123.00	1	1
10	95.00	0	26.0	113.00	1	1
11	96.00	0	19.0	95.00	3	0
12	97.00	0	19.0	150.00	3	0
13	98.00	0	22.0	95.00	3	0
14	99.00	0	30.0	107.00	3	0
15	100.00	0	18.0	100.00	1	1
16	101.00	0	18.0	100.00	1	1
17	102.00	0	15.0	98.00	2	0
18	103.00	0	25.0	118.00	1	1
19	104.00	0	20.0	120.00	3	0
20	105.00	0	28.0	120.00	1	1

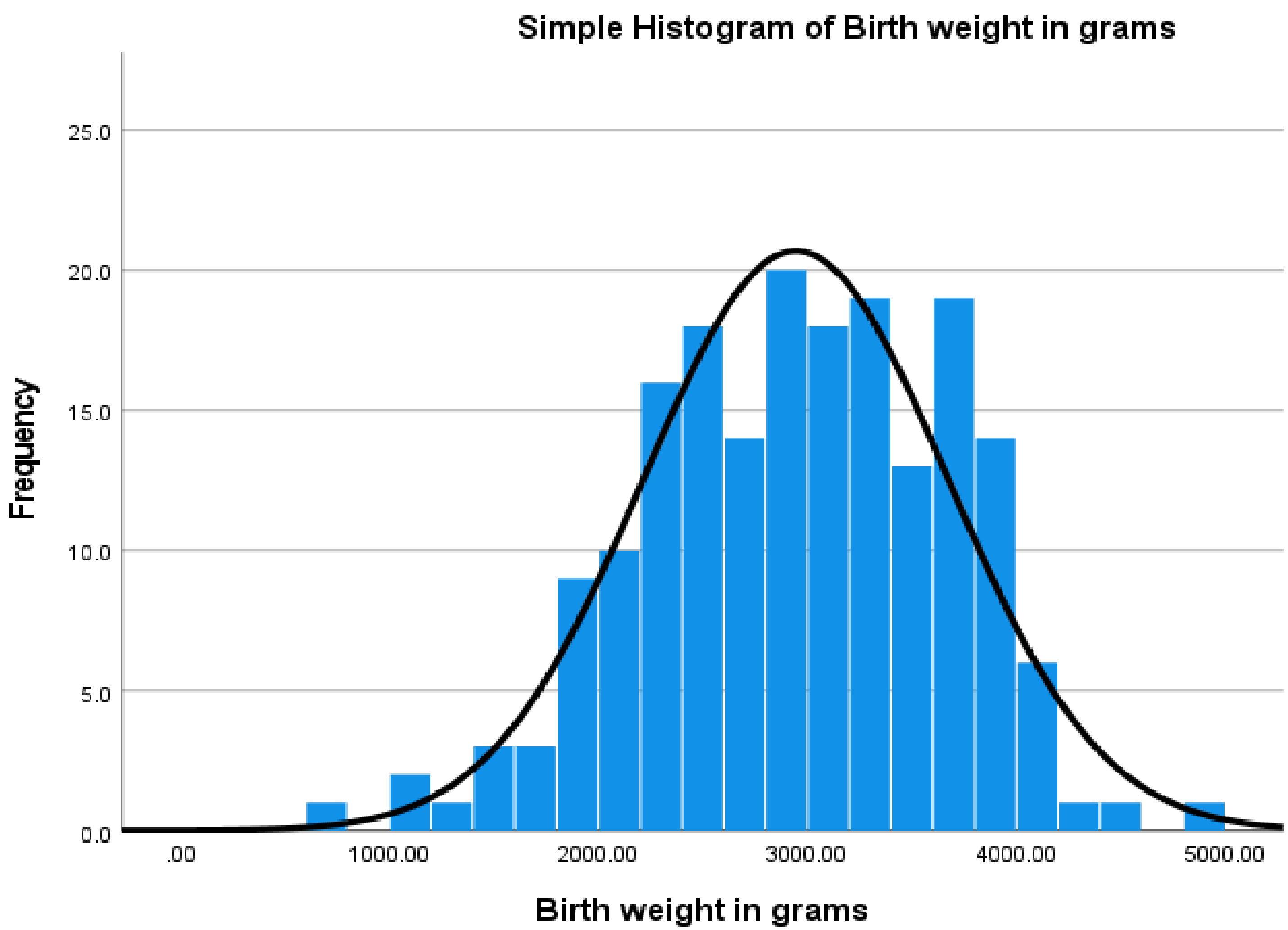
birthwt.sav

PREM	HYPER	URIN_IRR	PHYS_VIS	BIRTH_WT
0	0	1	0	2523.00
0	0	0	3	2551.00
0	0	0	1	2557.00
0	0	1	2	2594.00
0	0	1	0	2600.00
0	0	0	0	2622.00
0	0	0	1	2637.00
0	0	0	1	2637.00
0	0	0	1	2663.00
0	0	0	0	2665.00
0	0	0	0	2722.00
0	0	0	1	2733.00
0	1	0	0	2750.00
1	0	1	2	2750.00
0	0	0	0	2769.00
0	0	0	0	2769.00
0	0	0	0	2778.00
0	0	0	3	2782.00
0	0	1	0	2807.00
0	0	0	1	2821.00

Distribution of Birth Weight

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Birth weight in grams	189	4281.00	709.00	4990.00	2944.6561	53.02858	729.02242	531473.684	-.210	.177	-.081	.352
Valid N (listwise)	189											



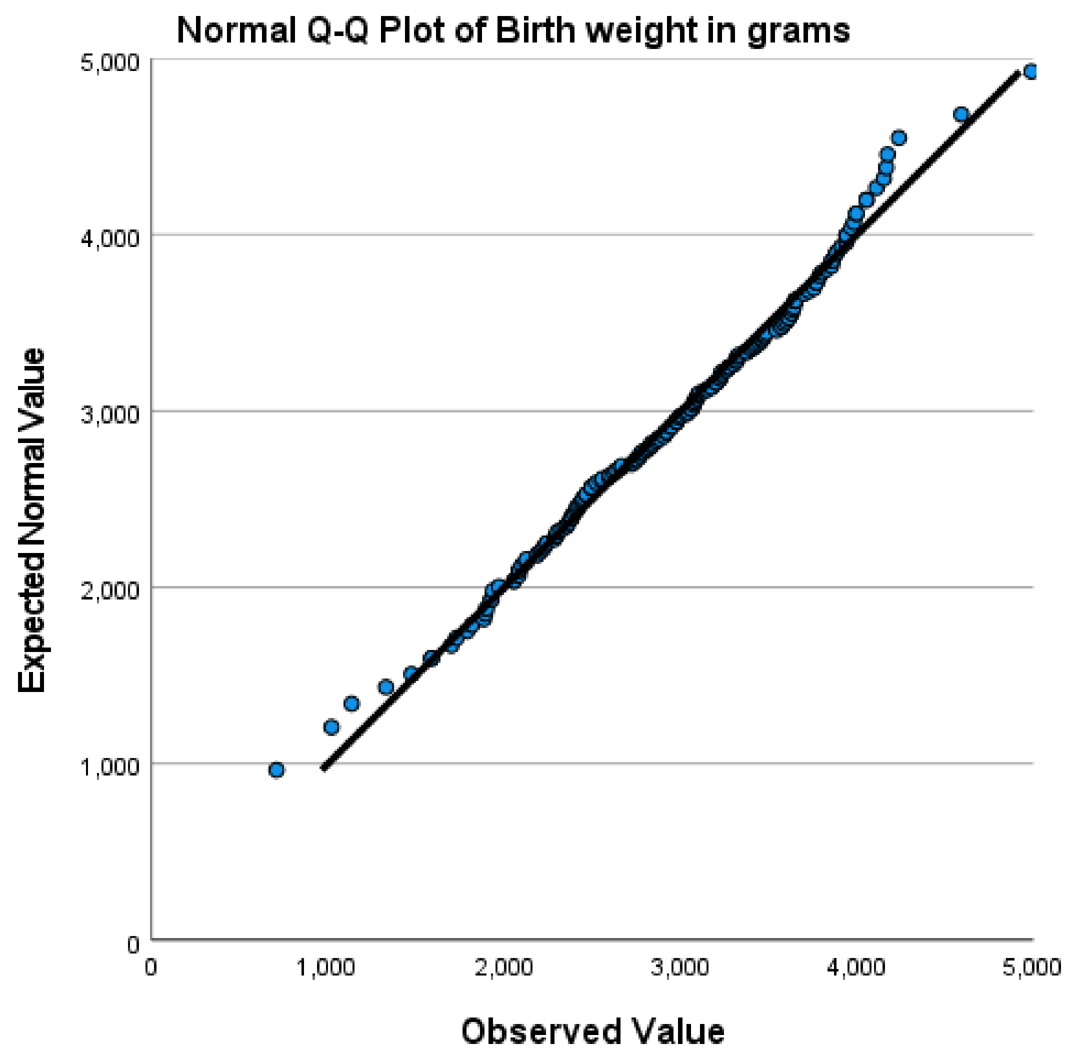
Mean = 2944.6561
Std. Dev. = 729.02242
N = 189

Analyze -> Descriptive Statistics-> Descriptives

<u>Options</u>
Mean, SD, median, range etc…

Normal Q-Q Plot of Baby's Birth Weight

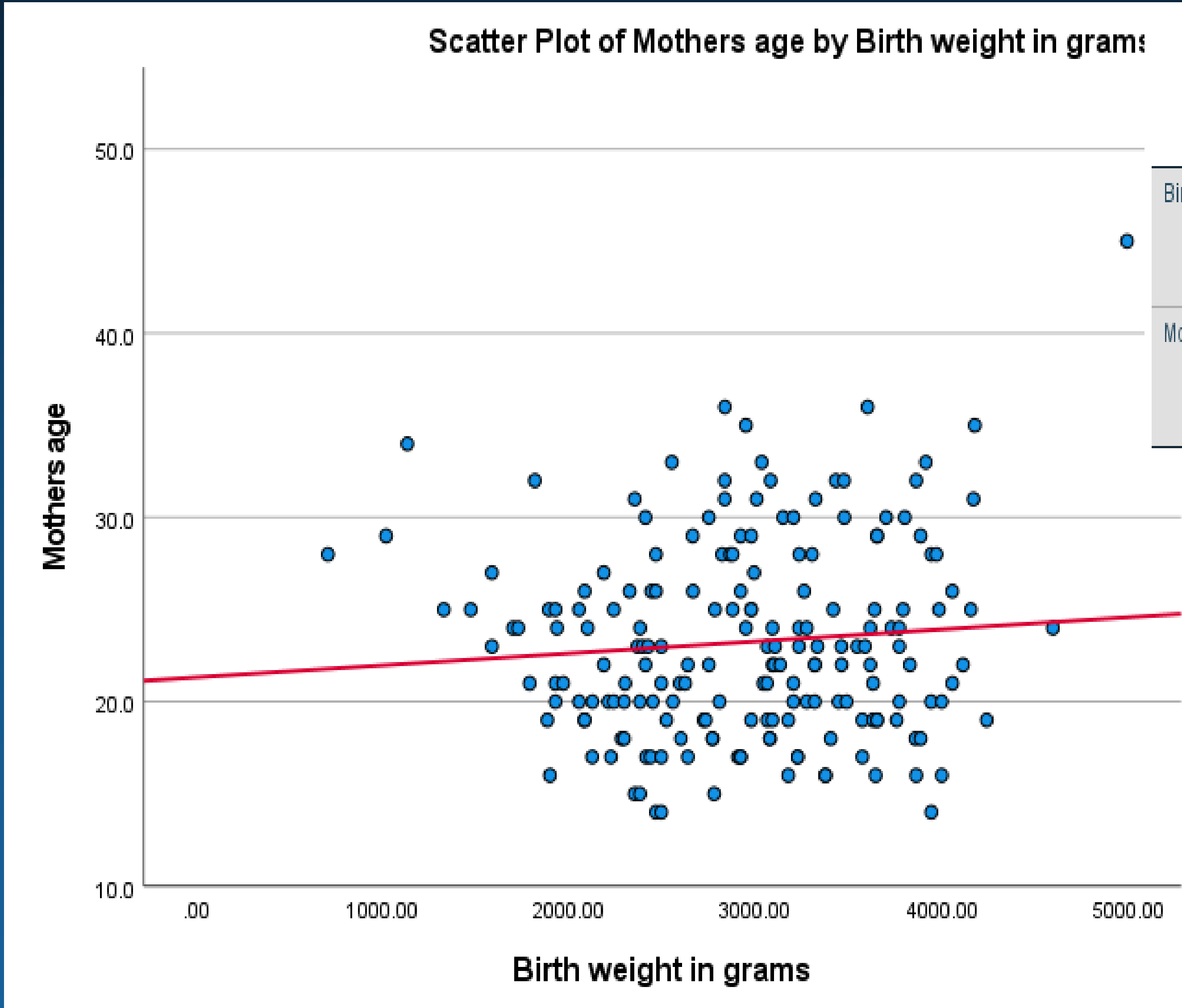
Analyze -> Descriptive Statistics-> Q-Q Plots



Correlation - Example

Is there a correlation between mother’s age and baby’s birthweight?

Analyze -> Correlate -> Bivariate

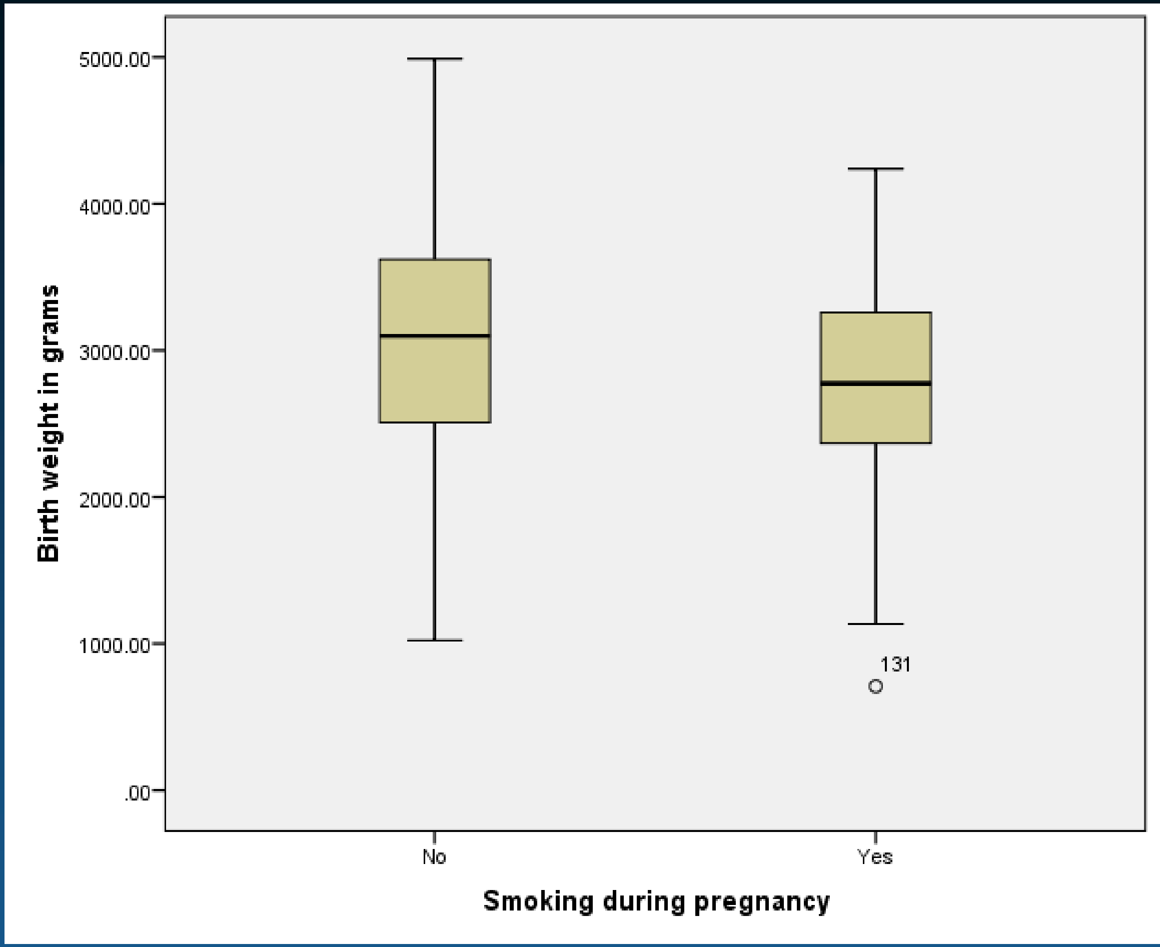


Correlations

		Birth weight in grams	Mothers age
Birth weight in grams	Pearson Correlation	1	.090
	Sig. (2-tailed)		.219
	N	189	189
Mothers age	Pearson Correlation	.090	1
	Sig. (2-tailed)	.219	
	N	189	189

Example:

Is there a statistically significant difference in baby's birth weight by maternal smoking during pregnancy?



Group Statistics

	Smoking during pregnancy	N	Mean	Std. Deviation	Std. Error Mean
Birth weight in grams	Yes	74	2773.2432	660.07517	76.73218
	No	115	3054.9565	752.40901	70.16250

T-Test - Example

Is there a statistically significant difference in baby's birth weight by mother smoking during pregnancy?

Analyze -> Compare Means -> Independent Samples T-Test

Group Statistics

	Smoking during pregnancy	N	Mean	Std. Deviation	Std. Error Mean
Birth weight in grams	Yes	74	2773.2432	660.07517	76.73218
	No	115	3054.9565	752.40901	70.16250

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Birth weight in grams	Equal variances assumed	1.508	.221	-2.634	187	.009	-281.71328	106.96873	-492.73382	-70.69274
	Equal variances not assumed			-2.709	170.001	.007	-281.71328	103.97406	-486.95979	-76.46677

ANOVA (Analysis of Variance)

Hypothesis: Is there a statistically significant difference in baby's birth weight by mother's race?

What if we want to compare means among 3 groups?

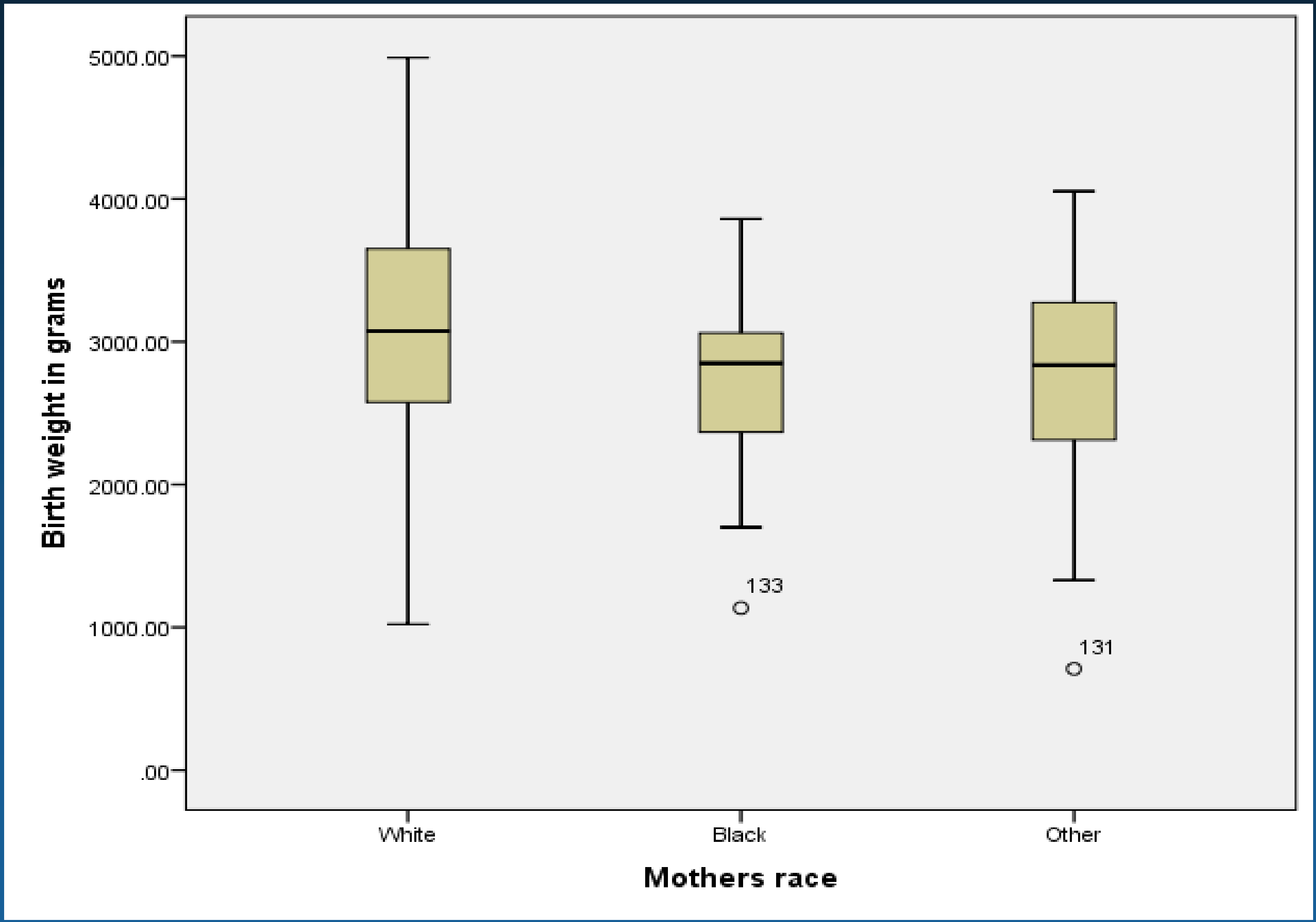
- Unfortunately the T test only allows us to compare two groups at a time: two sample T-test
- The T test is NOT appropriate for comparisons of 3 or more groups: issues with multiple comparisons

A global test that is used to compare the means of three or more groups

One way ANOVA: one independent variable

Anova- Example

Hypothesis: Is there a difference in baby's birth weight by mother's race?



Anova- Example

Analyze -> Compare Means -> One-Way ANOVA

Descriptives

Birth weight in grams

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
White	96	3103.7396	727.72424	74.27304	2956.2889	3251.1902	1021.00	4990.00
Black	26	2719.6923	638.68388	125.25621	2461.7223	2977.6623	1135.00	3860.00
Other	67	2804.0149	721.30115	88.12096	2628.0758	2979.9541	709.00	4054.00
Total	189	2944.6561	729.02242	53.02858	2840.0486	3049.2636	709.00	4990.00

ANOVA

Birth weight in grams

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5070607.632	2	2535303.816	4.972	.008
Within Groups	94846445.01	186	509927.124		
Total	99917052.65	188			

P is statistically significant, hence we reject H_0

At least one group mean is different from others

Post hoc Analysis – Race and Birthweight Which of the 3 groups are different?

Multiple Comparisons

Dependent Variable: Birth weight in grams
Scheffe

(I) Mothers race	(J) Mothers race	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
White	Black	384.04728	157.87439	.054	-5.5222	773.6168
	Other	299.72466*	113.67759	.033	19.2148	580.2345
Black	White	-384.04728	157.87439	.054	-773.6168	5.5222
	Other	-84.32262	164.99526	.878	-491.4635	322.8183
Other	White	-299.72466*	113.67759	.033	-580.2345	-19.2148
	Black	84.32262	164.99526	.878	-322.8183	491.4635

*. The mean difference is significant at the 0.05 level.

Chi-Square Test

Assume we wish to compare proportions of two birth weight groups by maternal hypertension during pregnancy

Birth weight group * History of hypertension Crosstabulation

			History of hypertension		Total
			No	Yes	
Birth weight group	>2500 gm	Count	125	5	130
		% within Birth weight group	96.2%	3.8%	100.0%
		% within History of hypertension	70.6%	41.7%	68.8%
	< 2500 gm	Count	52	7	59
		% within Birth weight group	88.1%	11.9%	100.0%
		% within History of hypertension	29.4%	58.3%	31.2%
Total		Count	177	12	189
		% within Birth weight group	93.7%	6.3%	100.0%
		% within History of hypertension	100.0%	100.0%	100.0%

$$X^2_{(df)} = \Sigma\ (Obs - Exp)^2 / Exp$$

Need to calculate expected values

Chi-Square Test

Analyze -> Descriptive Statistics-> Cross Tabulations

Birth weight group * History of hypertension Crosstabulation

			History of hypertension		Total
			No	Yes	
Birth weight group	>2500 gm	Count	125	5	130
		Expected Count	121.7	8.3	130.0
	< 2500 gm	Count	52	7	59
		Expected Count	55.3	3.7	59.0
Total		Count	177	12	189
		Expected Count	177.0	12.0	189.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.388[a]	1	.036		
Continuity Correction[b]	3.143	1	.076		
Likelihood Ratio	4.022	1	.045		
Fisher's Exact Test				.052	.042
Linear-by-Linear Association	4.365	1	.037		
N of Valid Cases	189				

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 3.75.

b. Computed only for a 2x2 table

Chi-Square Test
Can be used also for n x n tables

Birth weight group * Mothers race Crosstabulation

			Mothers race			
			White	Black	Other	Total
Birth weight group	>2500 gm	Count	73	15	42	130
		Expected Count	66.0	17.9	46.1	130.0
	< 2500 gm	Count	23	11	25	59
		Expected Count	30.0	8.1	20.9	59.0
Total		Count	96	26	67	189
		Expected Count	96.0	26.0	67.0	189.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	5.005[a]	2	.082
Likelihood Ratio	5.010	2	.082
Linear-by-Linear Association	3.570	1	.059
N of Valid Cases	189		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 8.12.

Linear Regression

Is there a relationship between baby birth weight and maternal hypertension during pregnancy, after adjusting for age and smoking?

Analyze -> Regression-> Linear

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6262001.401	3	2087333.800	4.123	.007[b]
	Residual	93655051.24	185	506243.520		
	Total	99917052.65	188			

a. Dependent Variable: Birth weight in grams

b. Predictors: (Constant), Smoking during pregnancy, History of hypertension, Mothers age

Model Summary[b]

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.250[a]	.063	.047	711.50792	.063	4.123	3	185	.007

a. Predictors: (Constant), Smoking during pregnancy, History of hypertension, Mothers age

b. Dependent Variable: Birth weight in grams

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	2824.666	239.603		11.789	.000	2351.960	3297.371
	History of hypertension	-424.465	212.287	-.142	-1.999	.047	-843.279	-5.651
	Mothers age	10.933	9.804	.079	1.115	.266	-8.409	30.276
	Smoking during pregnancy	-273.621	106.147	-.184	-2.578	.011	-483.035	-64.206

a. Dependent Variable: Birth weight in grams

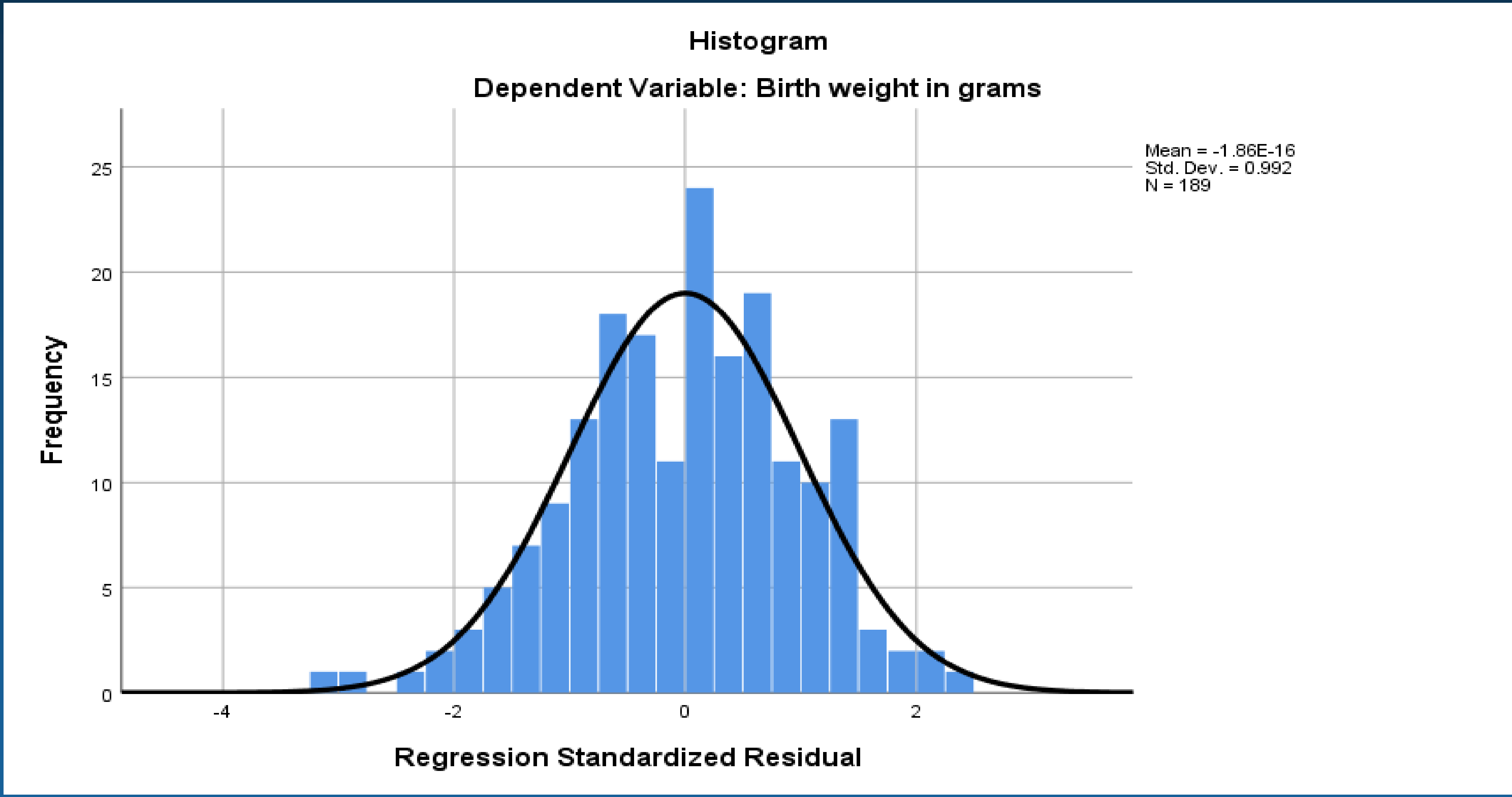
Analyze -> Regression-> Linear

Linear Regression (residuals)

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2334.3154	3316.6707	2944.6561	182.50621	189
Residual	-2148.18164	1673.32935	.00000	705.80817	189
Std. Predicted Value	-3.344	2.038	.000	1.000	189
Std. Residual	-3.019	2.352	.000	.992	189

a. Dependent Variable: Birth weight in grams



"To My Data, Right or Wrong"