


סיכום הקורס

פינגווינים!

Adelie

Chinstrap

Gentoo

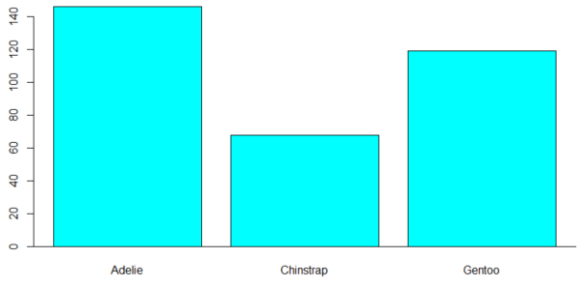
הנתונים – סך הכל 333 תצפיות

species	island	bill	weight	flipper	sex
Adelie	Torgersen	39.1	3750	short	male
Adelie	Torgersen	39.5	3800	short	female
Adelie	Torgersen	40.3	3250	medium	female
Adelie	Torgersen	36.7	3450	medium	female
Adelie	Torgersen	39.3	3650	medium	male
Adelie	Torgersen	38.9	3625	short	female
Adelie	Torgersen	39.2	4675	medium	male
Adelie	Torgersen	41.1	3200	short	female
Adelie	Torgersen	38.6	3800	medium	male
Adelie	Torgersen	34.6	4400	medium	male
Adelie	Torgersen	36.6	3700	short	female
Adelie	Torgersen	38.7	3450	medium	female
Adelie	Torgersen	42.5	4500	medium	male

הנתונים בשיעור זה הם נתונים מעובדים של הנתונים המקוריים:

Gorman KB, Williams TD, Fraser WR (2014) Ecological Sexual Dimorphism and Environmental Variability within a Community of Antarctic Penguins (Genus Pygoscelis). PLoS ONE 9(3): e90081. doi:10.1371/journal.pone.0090081

סטטיסטיקה תיאורית – משתנה בסולם שמי

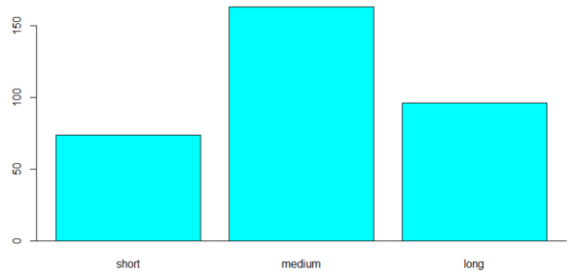


דיאגרמת עמודות

species	Freq	pct
Adelie	146	43.8%
Chinstrap	68	20.4%
Gentoo	119	35.7%
Total	333	99.9%

טבלת שכיחות

סטטיסטיקה תיאורית – משתנה בסולם סדר



flipper	Freq	cumm_freq	pct	cumm_pct
short	74	74	22.2%	22.2%
medium	163	237	48.9%	71.2%
long	96	333	28.8%	100%

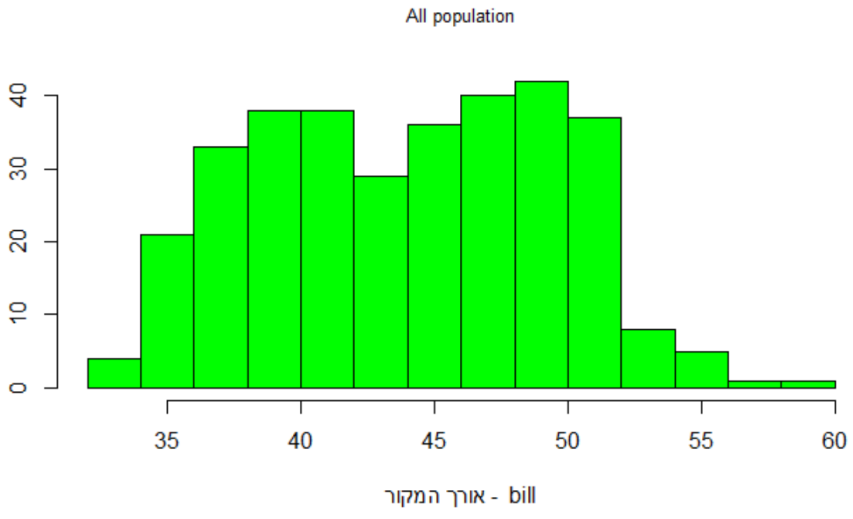
דיאגרמת עמודות

טבלת שכיחות ושכיחות מצטברת

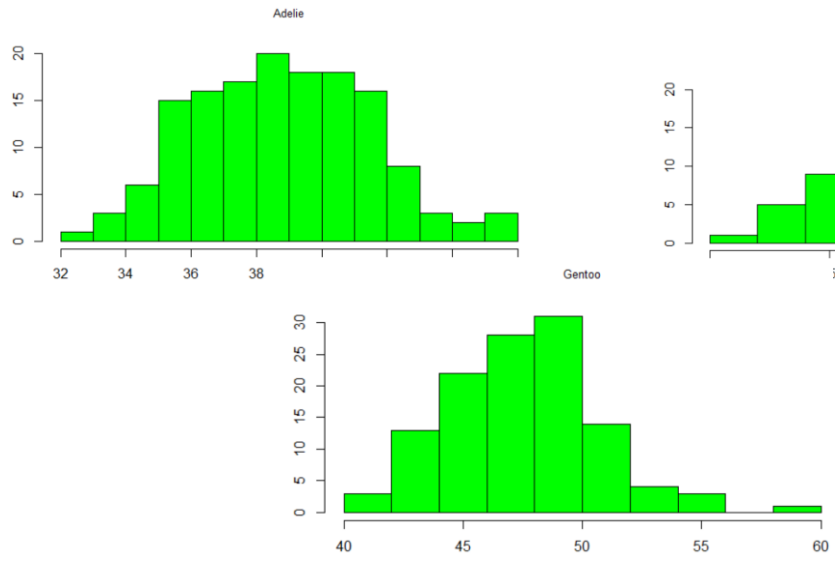
יש להקפיד על הסדר של הערכים

ניתן גם לחשב חציון, רבעונים וכדומה, אבל כאר מספר הערכים הוא קטן, אין לכך משמעות רבה

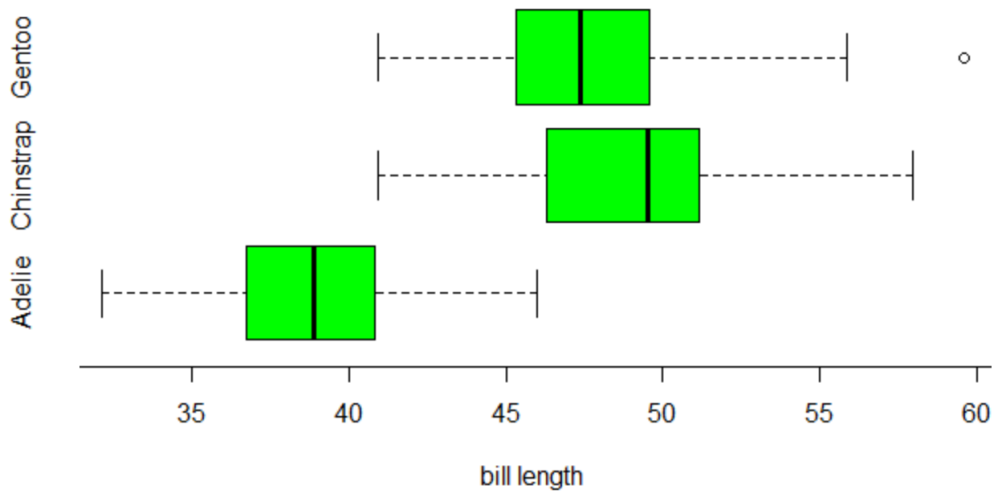
סטטיסטיקה תיאורית – משתנים כמותיים



אורך המקור – כל סוג פינגווין בנפרד



אורך המקור - boxplots



אורך המקור – סטטיסטיקה תיאורית

species	n	mean	sd	Q1	median	Q3
Adelie	146	38.8	2.7	36.7	38.8	40.8
Chinstrap	68	48.8	3.3	46.3	49.5	51.1
Gentoo	119	47.6	3.1	45.3	47.4	49.6

- מדדי מיקום מרכזי: ממוצע, חציון
- מדדי מיקום יחסי: רבעון תחתון, רבעון עליון
- מדדי פיזור: סטיית תקן, תחום בין רבעוני

אורך המקור, נתוני gentoo בלבד, n=119

```
  [1] 40.9 41.7 42.0 42.6 42.7 42.8 42.9 43.2 43.3 43.3
 [11] 43.4 43.5 43.5 43.6 43.8 44.0 44.4 44.5 44.9 44.9
 [21] 45.0 45.1 45.1 45.1 45.2 45.2 45.2 45.2 45.3 45.3
 [31] 45.4 45.5 45.5 45.5 45.5 45.7 45.8 45.8 46.1 46.1
 [41] 46.2 46.2 46.2 46.3 46.4 46.4 46.5 46.5 46.5 46.5
 [51] 46.6 46.7 46.8 46.8 46.8 46.9 47.2 47.2 47.3 47.4
 [61] 47.5 47.5 47.5 47.6 47.7 47.8 48.1 48.2 48.2 48.4
 [71] 48.4 48.4 48.5 48.5 48.6 48.7 48.7 48.7 48.8 49.0
 [81] 49.1 49.1 49.1 49.2 49.3 49.4 49.5 49.5 49.6 49.6
 [91] 49.8 49.8 49.9 50.0 50.0 50.0 50.0 50.1 50.2 50.4
[101] 50.4 50.5 50.5 50.5 50.7 50.8 50.8 51.1 51.1 51.3
[111] 51.5 52.1 52.2 52.5 53.4 54.3 55.1 55.9 59.6
```

מיקום החציון: $0.5 \cdot 119 = 59.5$

מיקום הרבעון התחתון: $0.25 \cdot 119 = 29.75$

מיקום הרבעון העליון: $0.75 \cdot 119 = 89.25$

אורך המקור, נתוני gentoo, זיהוי חריגים

```
  [1] 40.9 41.7 42.0 42.6 42.7 42.8 42.9 43.2 43.3 43.3
 [11] 43.4 43.5 43.5 43.6 43.8 44.0 44.4 44.5 44.9 44.9
 [21] 45.0 45.1 45.1 45.1 45.2 45.2 45.2 45.2 45.3 45.3
 [31] 45.4 45.5 45.5 45.5 45.5 45.7 45.8 45.8 46.1 46.1
 [41] 46.2 46.2 46.2 46.3 46.4 46.4 46.5 46.5 46.5 46.5
 [51] 46.6 46.7 46.8 46.8 46.8 46.9 47.2 47.2 47.3 47.4
 [61] 47.5 47.5 47.5 47.6 47.7 47.8 48.1 48.2 48.2 48.4
 [71] 48.4 48.4 48.5 48.5 48.6 48.7 48.7 48.7 48.8 49.0
 [81] 49.1 49.1 49.1 49.2 49.3 49.4 49.5 49.5 49.6 49.6
 [91] 49.8 49.8 49.9 50.0 50.0 50.0 50.0 50.1 50.2 50.4
[101] 50.4 50.5 50.5 50.5 50.7 50.8 50.8 51.1 51.1 51.3
[111] 51.5 52.1 52.2 52.5 53.4 54.3 55.1 55.9 59.6
```

$$IQR = Q_3 - Q_1 = 49.6 - 45.3 = 4.3$$

$$1.5 \cdot IQR = 1.5 \cdot 4.3 = 6.45$$

$$Q_1 - 1.5 \cdot IQR = 45.3 - 6.45 = 38.85$$

אין תצפיות חריגות כלפי מטה

$$Q_3 + 1.5 \cdot IQR = 49.6 + 6.45 = 56.05$$

תצפית אחת חריגה כלפי מעלה

אורך המקור, נתוני gentoo, זיהוי חריגים

```
  [1] -2.16 -1.90 -1.81 -1.61 -1.58 -1.55 -1.52 -1.42 -1.39 -1.39
 [11] -1.35 -1.32 -1.32 -1.29 -1.23 -1.16 -1.03 -1.00 -0.87 -0.87
 [21] -0.84 -0.81 -0.81 -0.81 -0.77 -0.77 -0.77 -0.77 -0.74 -0.74
 [31] -0.71 -0.68 -0.68 -0.68 -0.68 -0.61 -0.58 -0.58 -0.48 -0.48
 [41] -0.45 -0.45 -0.45 -0.42 -0.39 -0.39 -0.35 -0.35 -0.35 -0.35
 [51] -0.32 -0.29 -0.26 -0.26 -0.26 -0.23 -0.13 -0.13 -0.10 -0.06
 [61] -0.03 -0.03 -0.03  0.00  0.03  0.06  0.16  0.19  0.19  0.26
 [71]  0.26  0.26  0.29  0.29  0.32  0.35  0.35  0.35  0.39  0.45
 [81]  0.48  0.48  0.48  0.52  0.55  0.58  0.61  0.61  0.65  0.65
 [91]  0.71  0.71  0.74  0.77  0.77  0.77  0.77  0.81  0.84  0.90
[101]  0.90  0.94  0.94  0.94  1.00  1.03  1.03  1.13  1.13  1.19
[111]  1.26  1.45  1.48  1.58  1.87  2.16  2.42  2.68  3.87
```

$$\text{z-score} = \frac{x - mean}{sd}$$

$mean = 47.6$ $sd = 3.1$

רווחי סמך לתוחלות של אורכי המקורים

species	n	mean	sd	Q1	median	Q3
Adelie	146	38.8	2.7	36.7	38.8	40.8
Chinstrap	68	48.8	3.3	46.3	49.5	51.1
Gentoo	119	47.6	3.1	45.3	47.4	49.6

הצורה הכללית של רווח סמך לתוחלת: $\bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$$\text{Adelie: } 38.8 \pm 1.96 \cdot \frac{2.7}{\sqrt{146}} = (38.4, 39.2)$$

$$\text{Chinstrap: } 48.8 \pm 1.96 \cdot \frac{3.3}{\sqrt{68}} = (48.0, 49.6)$$

$$\text{Gentoo: } 47.6 \pm 1.96 \cdot \frac{3.1}{\sqrt{119}} = (47.0, 48.2)$$

רווח סמך לפרופורציה של פינגווני gentoo באוכלוסייה

הצורה הכללית של רווח סמך לפרופורציה: $\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

מתוך 333 פינגווינים בסך הכל, יש 119 פינגווני gentoo , ולכן $\hat{p} = \frac{119}{333} = 0.357$

לכן רווח הסמך לפרופורציה של פינגווני gentoo באוכלוסייה הוא

$$0.357 \pm 1.96 \cdot \sqrt{\frac{0.357 \cdot 0.643}{333}} = (0.306, 0.409)$$

בדיקת השערות –תוחלת אורך המקור של פינגווני Adelie

לצורך הדוגמה נניח כי סטיית התקן של ארכי המקורים ידועה ושווה ל-2.5

$$H_0: \mu_A = 38 \quad H_1: \mu_A > 38$$

נדחה את השערת האפס אם ממוצע המדגם גדול מ-38 באופן מובהק, כלומר כאשר

$$\bar{X} > \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}} = 38 + 1.645 \cdot \frac{2.5}{\sqrt{146}} = 39.1$$

ממוצע המדגם היה 38.8 ולכן לא נדחה את השערת האפס

בדיקת השערות –תוחלת אורך המקור של פינגווני Adelie

לצורך הדוגמה נניח כי סטיית התקן של ארכי המקורים ידועה ושווה ל-2.5

$$H_0: \mu_A = 38 \quad H_1: \mu_A \neq 38$$

לבדיקת השערה דו צדדית, יותר קל לחשב את רווח הסמך לתוחלת (או לפרופורציה):

$$\bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 38.8 \pm 1.96 \cdot \frac{2.5}{\sqrt{146}} = (38.4, 39.2)$$

38 אינו נמצא בתוך רווח הסמך ולכן נדחה את השערת האפס

בדיקת השערות –הפרשי תוחלות

לצורך הדוגמה נניח כי סטיות התקן של אורכי המקורים של פינגווני Chinstrap ו-Gentoo ידועות ושתיהן שוות ל-3.2

$$H_0: \mu_C = \mu_G \quad H_0: \mu_C > \mu_G$$

נדחה את השערת האפס כאשר

$$\bar{X}_C - \bar{X}_G > z_\alpha \cdot \sqrt{\frac{\sigma_G^2}{n_G} + \frac{\sigma_C^2}{n_C}} = 1.645 \cdot \sqrt{\frac{3.2^2}{119} + \frac{3.2^2}{68}} = 0.80$$

הפרש הממוצעים הוא $\bar{X}_C - \bar{X}_G = 48.8 - 47.6 = 1.2$ ולכן נדחה את השערת האפס

חישובי עצמה - תוחלת אורך המקור של פינגווני Adelie

הנחה: סטיית התקן ידועה ושווה ל-2.5

$$H_0: \mu_A = 38 \quad H_1: \mu_A > 38$$

כלל ההחלטה: דוחים את השערת האפס אם $\bar{X} > 39.1$

מהי עצמת המבחן אם התוחלת שווה-39.5?

$$P_{H_1}(\bar{X} > 39.1) = P\left(\frac{\bar{X} - 39.5}{2.5/\sqrt{146}} > \frac{39.1 - 39.5}{2.5/\sqrt{146}}\right) = P(Z > -1.93) = 0.973$$

חישובי גודל מדגם - תוחלת אורך המקור של פינגווני Adelie

הנחה: סטיית התקן ידועה ושווה ל-2.5

$H_0: \mu_A = 38 \quad H_1: \mu_A > 38$

כלל ההחלטה: דוחים את השערת האפס אם $\bar{X} > 39.1$

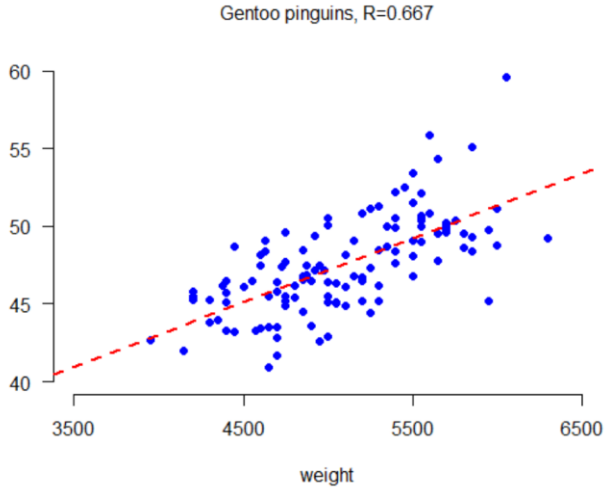
מה גודל המדגם הדרוש כדי לקבל עצמה של 80% אם התוחלת שווה ל-39.5?

$$P_{H_1}(\bar{X} > 39.1) = P\left(\frac{\bar{X} - 39.5}{2.5/\sqrt{n}} > \frac{39.1 - 39.5}{2.5/\sqrt{n}}\right) = P\left(Z > \frac{-0.4}{2.5/\sqrt{n}}\right) = 0.8$$

$$\frac{-0.4}{2.5/\sqrt{n}} = -1.282$$

$$\sqrt{n} = \frac{-1.282 \cdot 2.5}{-0.4} = 8.01 \qquad n = 65$$

מתאם: משקל ואורך מקור – פינגווני Gentoo



רגרסיה לינארית: משקל ואורך מקור – פינגווני Gentoo

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.667
R Square	0.445
Adjusted R Square	0.440
Standard Error	2.325
Observations	119

אחוז השונות המוסברת

סטיית התקן של השאריות

רגרסיה לינארית: משקל ואורך מקור – פינגווני Gentoo

ANOVA					
	df	*SS*	*MS*	*F*	*Sig F*
Regression		506.1	506.1	93.6	0.0000
Residual	117	632.4	5.4		
Total	118	1138.5			

$$SS_{Total} = SS_{Regression} + SS_{Residual}$$

$$1138.5 = 506.1 + 632.4$$

$$R^2 = SS_{Regression} / SS_{Total} = \frac{506.1}{1138.5} = 0.445$$

רגרסיה לינארית: משקל ואורך מקור – פינגווני Gentoo

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	26.538	2.184	12.2	0.0000	22.213	30.863
weight	0.004	0.000	9.7	0.0000	0.003	0.005

משתנים מסבירים

אמדני המקדמים

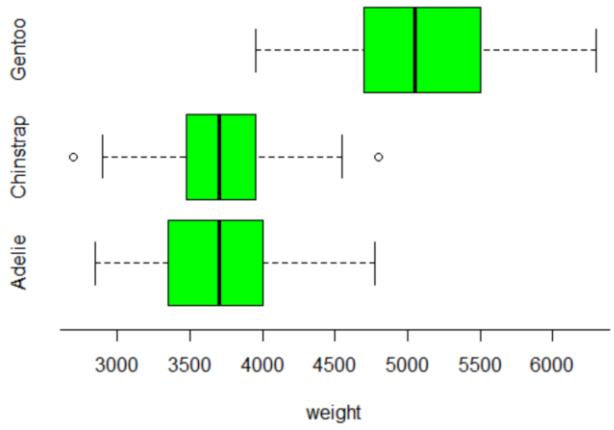
פי-ואליוז להשערות הדו-צדדיות כי המקדמים שווים לאפס

רווחי סמך לערכי המקדמים

משוואת הרגרסיה: $\text{bill} = 26.538 + 0.004 \cdot \text{weigth}$

ניתוח שונות: הבדלי משקל בין מיני הפינגווינים

$$H_0: \mu_A = \mu_C = \mu_G$$

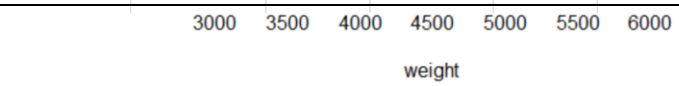


ניתוח שונות: הבדלי משקל בין מיני הפינגווינים

$$H_0: \mu_A = \mu_C = \mu_G$$

SUMMARY				
Groups	*Count*	*Sum*	*Average*	*Variance*
Adelie	146	541100	3706.2	210332.4
Chinstrap	68	253850	3733.1	147713.5
Gentoo	119	606000	5092.4	251478.3

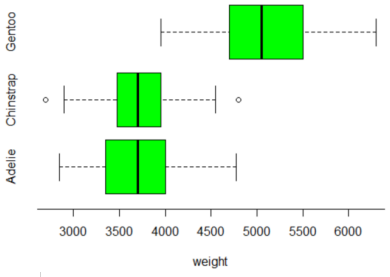
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	145190219.1	2	72595109.6	341.9	0.0000	3.0
Within Groups	70069446.8	330	212331.7			
Total	215259665.9	332				



הבדלי משקל בין מיני הפינגווינים: פוט הוק

SUMMARY				
Groups	*Count*	*Sum*	*Average*	*Variance*
Adelie	146	541100	3706.2	210332.4
Chinstrap	68	253850	3733.1	147713.5

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F cr*
Between Groups	33629.7	1	33629.7	0.2	0.6748	
Within Groups	40395003.5	212	190542.5			
Total	40428633	213				



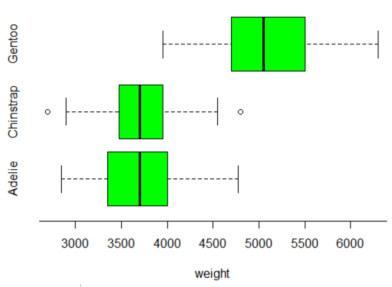
הבדלי משקל בין מיני הפינגווינים: פוט הוק

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Adelie	146	541100	3706.2	210332.4
Gentoo	119	606000	5092.4	251478.3

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	125994392.4	1	125994392.4	550.6908	0.0000
Within Groups	60172645.3	263	228793.3		
Total	186167037.7	264			



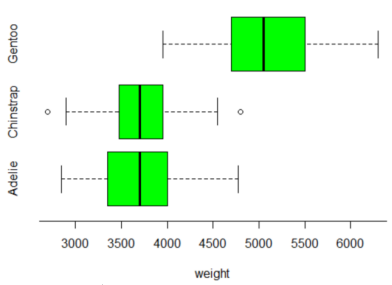
הבדלי משקל בין מיני הפינגווינים: פוט הוק

SUMMARY

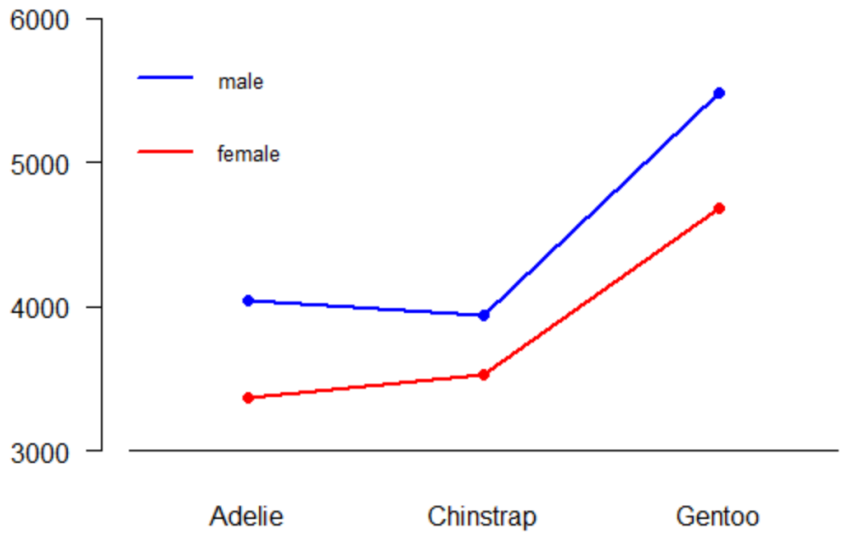
Groups	*Count*	*Sum*	*Average*	*Variance*
Chinstrap	68	253850	3733.1	147713.5
Gentoo	119	606000	5092.4	251478.3

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	79960600.2	1	79960600.2	373.8	0.0000
Within Groups	39571244.7	185	213898.6		
Total	119531844.9	186			



ניתוח שונות דו כיווני: משקל לפי סוג ומין



ניתוח שונות דו כיווני: משקל לפי סוג ומין

	ANOVA						
	Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
מין	Sample	24163347.2	1.0	24163347.2	243.9	0.0000	3.9
סוג הפינגווין	Columns	70446006.9	2.0	35223003.5	355.5	0.0000	3.0
אינטראקציה	Interaction	242840.3	2.0	121420.1	1.2	0.2961	3.0
	Within	17239041.7	174.0	99075.0			
	Total	112091236.1	179				

לוחות שכיחות – פיזור האוכלוסייה על פני האיים

	Biscoe	Dream	Torgersen
Adelie	44	55	47
Chinstrap	0	68	0
Gentoo	119	0	0

ברור לחלוטין כי יש קשר בין סוגי הפינגווינים ובין האיים בהם הם מתגוררים:

פינגווני Gentoo חיים רק באי Biscoe
פינגווני Cinstrap חיים רק באי Dream
באי Torgersen חיים רק פינגווני Adelie

פיזור פינגווני Adelie על פני האיים

	Biscoe	Dream	Torgersen
Adelie	44	55	47

פינגווני Adelie חיים בכל שלושת האיים. האם הם מתפזרים על פני האיים בצורה שווה?

	Biscoe	Dream	Torgersen	Total
Observed	44	55	47	146
probabilities	1/3	1/3	1/3	
Expected	48.67	48.67	48.67	
$(O-E)^2/E$	0.447	0.824	0.057	1.328

$\chi^2 = 1.328 \quad df = 2 \quad \text{p-value} = 0.5146$

פינגווינים Adelie: הקשר בין אורך הכנף והמין

האם יש קשר בין אורך הכנף והמין בקרב פינגווינים Adelie?

Observed	short	medium
female	43	30
male	19	54

Expected	short	medium
female	31	42
male	31	42

$(O-E)^2/E$	short	medium
female	4.645	3.423
male	4.645	3.423

$$\chi^2 = 16.15 \quad \text{df} = 1 \quad \text{p-value} = 0.00012$$