

שונות		תוחלת	פונקציית הסתברות	התפלגות
pq		p	פיימ - $Mx(t) = q + e^t$	<b>ברנולי</b> : ניסוי שהסתברות להצלחה היא p ולכישלון q
$n \cdot p \cdot q$		$n \cdot p$	פיימ: $Mx(t) = (q + pe^t)^n$ $P(X=k) = \binom{n}{k} p^k q^{n-k} X \sim Bin(n, p)$	<b>בינומית</b> - סידרה של n ניסויי ברנולי בלתי תלויים, עם הסתברות הצלחה p, מודד את מספר ההצלחות
$q/p^2$		$1/p$	פיימ: $Mx(t) = e^t p(1 - e^t q)^{-1}$ $P(X=k) = q^{k-1} p X \sim Geo(p)$	<b>גיאומטרית</b> - סידרה של n ניסויי ברנולי בלתי תלויים עם הסתברות הצלחה p, מודד את מספר הניסויים עד ההצלחה הראשונה
$\lambda$	$\lambda$		$P(X=k) = \frac{e^{-\lambda} \lambda^k}{k!} X \sim Pois(\lambda)$	<b>פואסונית</b> - מודד את מספר המכוניות שעברו עד זמן k כאשר המכוניות באות בקצב של $\lambda$ מכוניות ליחידת זמן בממוצע
$Cov(x_i, x_j) = \text{inp}_i p_j$	$E(x_i) = n \cdot p_i$	$P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k} X \sim Mult(n, p_1, p_2, \dots, p_k)$		<b>מולטינומית</b> - מבצעים n ניסויים בלתי תלויים שלכל אחד מהם יתכנו k תוצאות אפשריות עם הסתברות $p_1, p_2, \dots, p_k$ מודד את מספר התוצאות מכל סוג
$(X) = \frac{(b-a)^2}{12}$	$E(X) = \frac{b+a}{2}$	$Mx(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$ פיימ: $p(c \leq x \leq d) = \frac{d-c}{b-a}$	$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & otherwise \end{cases}$	<b>יוניפורמית</b> - בחירה של נקי אקראית בין a ל-b $F_X(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x \geq b \end{cases}$
$Var(X) = \frac{1}{\lambda^2}$	$E(X) = \frac{1}{\lambda}$ $E(X^2) = \frac{2}{\lambda^2}$	$y = \min(x_1 \dots x_n), x_i \sim \exp(\lambda)$ $y \sim \exp(n\lambda)$	$Mx(t) = \frac{\lambda}{\lambda - t}$ פיימ: $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & אחרת \end{cases}$	$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$ <b>אקספוננציאלית</b>
$F_z(a) = P(Z \leq a) = \Phi(a)$ $P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a) = \Phi(b) - \Phi(a)$ $\Phi(-a) = 1 - \Phi(a)$ $P(-a \leq Z \leq a) = 2\Phi(a) - 1$	חישוב הסתברויות	$E(X) = \mu$ $Var(X) = \sigma^2$	$z = \frac{x - \mu}{\sigma} \sim N(0,1)$ :זנא $x \sim N(\mu, \sigma^2)$ :זנא $f_X(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ $P(X \leq b) = P\left(\frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = P\left(Z \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right)$	<b>נורמלית</b> $X \sim N(\mu, \sigma^2)$
$Var(X) = \frac{n}{\lambda^2}$	$E(X) = \frac{n}{\lambda}$	$P(Z < z) = 1 - \sum_{k=0}^{n-1} \frac{e^{-\lambda z} (\lambda z)^k}{k!}$	$f_X(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{n-1}}{\Gamma(n)} & x \geq 0 \\ 0 & otherwise \end{cases}$	<b>Gamma</b> $\sim Gamma(n, \lambda)$

**נוסחאות חשובות:**

<p><b>הגדרות:</b> פרמטר - קבוע המאפיין את האוכלוסייה. לדוגמה: תוחלת, שונות, מקסי. * סטטיסטי - ביטוי כמותי הניתן לחישוב כפונקציה של התצפיות (ממוצע, סטיית תקן). - תלוי בדברים ידועים בלבד. * אמד - סטטיסטי שבעזרתו נאמוד פרמטר של התפלגות. * אומדן - הערך המיספרי של האמד עבור מדגם נתון.</p>	$Cov(x_1, x_2) = 0 \Leftrightarrow X_1, X_2 \text{ ת"ב}$ $Cov(x_1, x_2) = E(x_1 \cdot x_2) - E(x_1) \cdot E(x_2)$	$Var(X) = E(X^2) - [E(X)]^2$ $E((x - E(x))^2) = Var(x)$ $Var(\bar{x}) = \frac{\sigma^2}{n}$ $f_x(x) = \frac{dF(x)}{dx}, p(X < x) = F(x)$	$z = \max(x_1, \dots, x_n) \Rightarrow f_z(z) = n[F_x(x)]^{n-1} \cdot f_x(x)$ <b>התפלגות של מקסימום של מדגם</b>
			<b>התפלגות של מינימום:</b> $f_z(z) = n[1 - F_x(x)]^{n-1} \cdot f_x(x)$
			<b>תוחלת של אינדיקטור:</b> $E(I_A) = E(I_A^2) = P(A)$

**קריטריונים לבחירת אמד:**

**אמד חסר הטייה:**

אמד  $\hat{\Theta}$  ל- $\Theta$  יקרא אמד חסר הטייה אם  $E(\hat{\Theta}) = \Theta$

• א.ח.ה ל- $\mu$  (תוחלת):  $\hat{\mu} = \bar{X}$

• א.ח.ה ל- $\sigma^2 = \frac{\sum (Xi - \bar{X})^2}{n-1}$  אם התוחלת נתונה:  $\sigma^2 = \frac{\sum (Xi - \bar{X})^2}{n}$

• בחירה בין 2 א.ח.ה: אמד  $\hat{\Theta}_1$  עדיף על- $\hat{\Theta}_2$  (כאשר שניהם א.ח.ה) אם  $Var(\hat{\Theta}_1) \leq Var(\hat{\Theta}_2)$

• אח"ה למ"מ מעריכי:  $x_i \sim \exp(\lambda)$  מומנט אוכלוסיה:  $\mu = e(x) = 1/\lambda$ , אמד של מומנט:

$$\frac{1}{\hat{\lambda}} = \bar{x} \Rightarrow \hat{\lambda} = \frac{1}{\bar{x}} \cdot \hat{\mu} = \frac{\sum x_i}{n} = \bar{x}$$

**שיטות אמידה:**

**שיטת המומנטים:**

• מומנט של אוכלוסייה: המומנט ה- $K$  של  $X$  מסומן  $\mu_x$ :  $\mu_k = E(x^k)$

• מומנטים מדגמיים: המומנט ה- $K$  של המדגם מסומן ב- $\hat{\mu}_k$ :  $\hat{\mu}_k = \frac{\sum x_i^k}{n}$

•  $\hat{\mu}_k$  הוא א.ח.ה ל- $\mu_x$ .

• שלבים למציאת אמד בשיטת המומנטים:

1. מציאת מומנט האוכלוסייה.
2. מציאת אמד למומנט האוכלוסייה (מומנט המדגם).
3. מומנט האוכלוסייה = אמד למומנט האוכלוסייה.
4. חילוץ האמד.

**אמדים בשיטת המומנטים:**

**M.S.E:**

• בהינתן  $\hat{\Theta}$  אמד ל- $\Theta$ :  $MSE_{\hat{\Theta}}(\Theta) = [E(\hat{\Theta} - \Theta)]^2$

• נוסחא חישובית:  $MSE_{\hat{\Theta}}(\Theta) = Var(\hat{\Theta}) + [Bias_{\hat{\Theta}}(\hat{\Theta})]^2$

•  $Bias_{\hat{\Theta}}(\hat{\Theta}) = E(\hat{\Theta}) - \Theta$  אם  $\hat{\Theta}$  א.ח.ה  $= 0$

• קריטריון: נעדיף אמד  $\hat{\Theta}_1$  על  $\hat{\Theta}_2$  אם:

$$MSE(\hat{\Theta}_1) \leq MSE(\hat{\Theta}_2)$$

**אמד נראות מקסימלי (א.נ.מ.):**

1. נמצא  $f_{\Theta}(x_i)$

2. נכתוב:  $L(\Theta) = f_{\Theta}(x_1, \dots, x_n) = \prod f_{\Theta}(x_i)$

3. במקרה ונח יותר לחשב נעבור ל:  $\ln(L(\Theta))$

•  $\ln(x^\alpha) = \alpha \ln(x)$ ,  $\ln(e^x) = x$ ,  $\ln(\prod) = \sum(\ln)$

4. נמצא מקסימום על ידי גזירה לפי  $\Theta$  והשוואה ל-0.

5. נוודא שהנקודה היא מקסי' על ידי גזירה שנייה קטנה מ-0.

**אמדים בשיטת א.נ.מ.:**

<p><b>בדיקת השערות חד צדדיות עבור מ"מ נורמלי:</b></p> $z \sim N(0,1), z = \frac{\mu - x}{\sigma} \Leftarrow x \sim N(\mu, \sigma^2) \quad .1$ $z_n = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \sim N(0,1), \Leftarrow x_1 \dots x_n \sim N(\mu, \sigma^2) \quad .2$ $z_n = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \rightarrow N(0,1), \Leftarrow x_1 \dots x_n \sim F \quad .3$	<p><b>סוגי טעויות (עבור השערות פשוטות):</b></p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td></td> <td><b>H1 הנוכח</b></td> <td><b>H0 הנוכח</b></td> </tr> <tr> <td><b>טעות מסוג 1</b></td> <td>עוצמה</td> <td>דחיית H0</td> </tr> <tr> <td><b>רמת ביטחון</b></td> <td><b>טעות מסוג 2</b></td> <td>קבלת H0</td> </tr> </table> <p>שגיאה מסוג 1: <math>P(R   H0) = P_{H0}(R)</math>  שגיאה מסוג 2: <math>P(\bar{R}   H1) = P_{H1}(\bar{R})</math></p>		<b>H1 הנוכח</b>	<b>H0 הנוכח</b>	<b>טעות מסוג 1</b>	עוצמה	דחיית H0	<b>רמת ביטחון</b>	<b>טעות מסוג 2</b>	קבלת H0	<p><b>פונקציה עוצמה</b> - <math>(\pi(x))</math> - ההסתברות תחת H0 לחדות את H0 - <math>P_{H0}(R)</math>.</p> <p><math>\alpha</math> - רמת מובהקות: ערך המביא את פונקציה העוצמה למקסימום (= ההסתברות לטעות מסוג 1)  * בהינתן 2 גישות לבדיקת השערות נעדיף את זו עברה פונקציה העוצמה גדולה יותר.</p>	<p><b>בדיקת השערות:</b> קיימות 2 השערות H0: (השערת האפס) ו-H1: (השערה אלטרנטיבית). קיימות 2 החלטות אפשריות: 1. דחיית H0. 2. אי דחיית H0. ההחלטה נעשית ע"י כלל החלטה: נקבע איזור דחייה R - תחום בו מחליטים לחדות את H0 והאזור המשלים אותו הוא איזור הקבלה.</p>
	<b>H1 הנוכח</b>	<b>H0 הנוכח</b>										
<b>טעות מסוג 1</b>	עוצמה	דחיית H0										
<b>רמת ביטחון</b>	<b>טעות מסוג 2</b>	קבלת H0										
<p><b>בדיקת השערות על תוחלת ללא הנחת נורמלי ושונות לא ידועה:</b></p> <p>1. מקרה חד צדדי: <math>x_1, \dots, x_n \sim F</math></p> $R = \left\{ \frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{(1-\alpha)}^{n-1} \right\}$ (ע"פ הטבלה) <p>2. מקרה דו צדדי - (ע"פ הטבלה)</p>	<p><b>בדיקת השערות על ערך התוחלת של מ"מ נורמלי - שונות ידועה:</b></p> <p>1. חד צדדי: <math>H0: \theta \leq \theta_0, H1: \theta &gt; \theta_0</math></p> <p>(מבחן ע"פ טבלה) אם <math>\leq</math>: <math>\bar{X} &gt; \frac{\sigma \cdot Z_{1-\alpha}}{\sqrt{n}} + \theta_0</math> נדחה את H0</p> <p>פונקציה עוצמה: <math>\Pi(\theta) = 1 - \Phi\left(Z_{1-\alpha} + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma}\right)</math></p> <p>2. דו-צדדי: <math>H0: \theta = \theta_0, H1: \theta \neq \theta_0</math> (מבחן ע"פ הטבלה)</p>	<p><b>סימון:</b> <math>\phi</math> - פונקציה התפלגות מצטברת של נורמלי סטנדרטי</p> $Z_{1-\alpha} = \phi^{-1}(1-\alpha)$ <p><b>אמד עקיב:</b> <math>\hat{\theta}</math> אמד עקיב ל-<math>\theta</math> אם: <math>P_\theta( \hat{\theta} - \theta  &gt; \epsilon) \rightarrow 0</math></p> <p>אם S אמד עקיב ל-<math>\sigma</math> אזי: <math>\frac{\sqrt{n}(x - \mu)}{s} \xrightarrow{n \rightarrow \infty} N(0,1)</math></p>										
<p><math>X1 \sim pois(\lambda1), X2 \sim pois(\lambda2) \Leftarrow</math> קיימות 2 השערות: <math>H0: \lambda1 = \lambda2</math>  <math>\Leftarrow H1: \lambda1 &lt; \lambda2</math></p> $\frac{X1 - X2 - (\lambda1 - \lambda2)}{\sqrt{x1 + x2}} \sim N(\lambda1 - \lambda2, \lambda1 + \lambda2)$	<p><b>מצייאת פונקציית עוצמה בהינתן איזור דחייה R:</b></p> <p><math>\Pi(p) = Pp(R)</math> נרצה להגיע לאחד הביטויים מהצורה:</p> <div style="border: 1px solid black; padding: 5px; width: fit-content;"> <p>בדיק ע"י חיבור וחסור שח המשתנה עליו מתבססות ההנחות והפרדה ל- גורמים</p> </div> $P_p\left(\frac{\sqrt{n}(x - \mu)}{\sigma} > c\right) = 1 - \phi(c)$ $P_p\left(\frac{\sum Xi - n\mu}{\sigma \cdot \sqrt{n}} < c\right) = \phi(c)$	<p><b>קביעת גודל המדגם:</b> <math>x_1, \dots, x_n \sim F</math></p> <p>ב"ת ו"ש"ה. עבוד מבחן חד צדדי מצא N גדול מספיק כך שיתקיים:</p> $n = \left[ \frac{(Z_{\pi_0} + Z_{1-\alpha}) \cdot \sigma}{\mu_1 - \mu_0} \right]^2$ <p><math>\pi(\mu_1) = \pi_0, \pi(\mu_0) = \alpha</math></p>										
<p><b>השוואת תוחלות של 2 אוכלוסיות:</b> נתונות 2 אוכלוסיות שונות:</p> <p>1. דגימה לא מזווגת (האוכלוסיות ב"ת האחת בשנייה): (ע"פ המקרים השונים בטבלה) פונקציית עוצמה:</p> $\Pi \approx 1 - \Phi\left(Z_{1-\alpha} + \frac{(\mu_2 - \mu_1)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right)$ <p>2. דגימה מזווגת (האוכלוסיות תלויות האחת בשנייה) נגדיר: <math>v_i = x_i - y_i, cov(x_i, y_i) = a &gt; 0</math></p> $E(v_i) = \mu_1 - \mu_2, var(v_i) = \sigma_1^2 + \sigma_2^2 - 2a$ <p>H0: <math>E(v_i) \leq 0 \Leftarrow H0: \mu_1 \leq \mu_2</math>  H0: <math>E(v_i) &gt; 0 \Leftarrow H0: \mu_1 &gt; \mu_2</math></p> <p>פונקציית עוצמה:</p> $\Pi \approx 1 - \Phi\left(Z_{1-\alpha} + \frac{(\mu_2 - \mu_1)\sqrt{n}}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2a}}\right)$ $R = \left\{ (x, y) \mid \frac{\sqrt{n}(\bar{x} - \bar{y})}{\sqrt{\frac{\sum (v_i - \bar{v})^2}{n}}} > Z_{1-\alpha} \right\}$	<p><b>שלבים במתירת בעיית בדיקת השערות:</b></p> <ol style="list-style-type: none"> <li>ניסוח השערות וקביעת נתונים (S)</li> <li>מצייאת סטטיסטי המבחן: ע"פ H0 והנתונים בודקים בטבלה מהו סטטיסטי המבחן (עמודה 2). <math>\leq</math> מצייאת P value (ע"פ הטבלה המתאימה לסטטיסטי המבחן נמצא את <math>(1-\alpha)</math>)</li> <li>מצייאת הערך הקריטי - ע"פ H1 והעמודה האחרונה של הטבלה נמצא את הערך הקריטי עבורו נדחה את H0. נקבל: <math>X &lt; C</math> כאשר X הוא סטטיסטי המבחן ו-C הוא הערך הקריטי</li> <li>ביצוע כלל ההחלטה.</li> <li>הסקת מסקנות (במילים)</li> </ol>	<p><b>P-Value:</b> בהינתן אוסף מבחנים ברמת מובהקות <math>0 &lt; \alpha \leq 1</math>, עם איזור דחייה <math>R_\alpha</math> נגדיר:</p> $P\_value(x) = \inf(\alpha \mid X \in R_\alpha)$ <p>(<math>\Leftarrow</math>) ההסתברות לקבל את התוצאה שקיבלנו במדגם או קיצונית ממנה תחת H0.</p> <p><math>\Leftarrow</math> בהינתן מבחן ברמת מובהקות <math>\alpha</math>: אם <math>P\_value &lt; \alpha</math> נדחה את H0</p> <p><math>\Leftarrow</math> בהינתן מבחן וערך סטטיסטי: מבחן דו-צדדי:</p> <p>ערך סטטיסטי <math>\leq Z_{1-\alpha/2} = P\_value = \alpha</math> מבחן חד צדדי:</p> <p>ערך סטטיסטי <math>\leq Z_{1-\alpha/2} = P\_value - \alpha</math> מקרה נורמלי:</p> <p><math>P\_value = P_{H0}(Z &gt; c) = 1 - \phi(c)</math></p>										

**רווח סמך / קבוצת סמך**

**הגדרה:**  $X \sim F_{\theta}$ ,  $\theta \in A(x)$  היא קבי סמך ל- $\theta$  ברמת סמך  $(1-\alpha) \cdot 100\%$  אם:  $P_{\theta}(\theta \in A(x)) = 1 - \alpha$   
 $A(x)$  – קבוצת סמך ברמת סמך  $1-\alpha$ .

• כאשר לקבוצת סמך צורת אינטרוול נאפיין את  $A(x) = (L(x), U(x))$ .

**רווח סמך לתוחלת:**  $x_1, \dots, x_n \sim F$  ביית  $E(X_i) = \theta$ . נרצה רווח סמך ברמת סמך  $(1-\alpha) \cdot 100\%$ .

$$l = 2 \cdot Z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}} - \bar{x}, \quad u = 2 \cdot Z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}} + \bar{x}, \quad \theta \in (\bar{x} - Z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + Z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}})$$

• על מנת לקבל אורך רווח סמך ברווח מסויים נחלץ את  $n$  מהביטוי המיל.

**דואליות בין רווח סמך ובדיקת השערות:**  $X \sim F_{\theta}$ , נניח לכל השערה דו"צ קיים מבחן ברמת מובהקות  $\alpha$  המבוסס על  $A(x) = \{\theta_0 \mid X \notin R(\theta_0)\} \in R$  היא קבוצת סמך ברמת סמך  $(1-\alpha) \cdot 100\%$ . (אותו הדבר הפוך).

• שלבים במציאת רווח סמך (עבור פרמטר  $\theta$ ):

1. בחירת אמד  $\hat{\theta}$  ל- $\theta$ .
2. בחירת פונקצייה  $Q(\theta, \hat{\theta})$  ע"פ הטבלה (עמודה 2).
3. חישוב שברונים של  $Q(X_{p1}, X_{p2})$  כך ש- $P2 - P1 = 1 - \alpha$ . (בהתאם להתפלגות של  $Q$ ).
4. כתיבת הדרישה בצורה הבאה:  $P(X_{p1} \leq Q \leq X_{p2}) = 1 - \alpha$
5. חילוץ  $\theta$  מהביטוי 4- וקבלת רווח סמך (לכתוב מילולית בסוף עבור איזה רמת סמך רווח הסמך התקבל).

**התפלגויות הקשורות למ"מ נורמליים:**

$$\sum_{i=1}^n z_i^2 = \sum_{i=1}^n \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2 \sim \chi^2(n), \quad x_1, \dots, x_n \sim N(\mu, \sigma^2) - \chi^2$$

$$\frac{\sqrt{n}(\bar{x} - \mu)}{s} \sim t(n-1), \quad x_1, \dots, x_n \sim N(\mu, \sigma^2), \quad \frac{Z}{\sqrt{v/k}} \sim t(k) \Leftrightarrow v \sim \chi^2(k), Z \sim N(0,1) : t$$

$$\frac{u/m}{v/k} \sim F(m, k) \Leftrightarrow v \sim \chi^2(k), u \sim \chi^2(m) : F$$

**נקודות חשובות לגבי רווח סמך:**

- רמת ביטחון גדלה  $\Leftrightarrow$  רווח סמך יגדל.
- מדגם גדול יותר  $\Leftrightarrow$  רווח סמך יגדל.
- שונות קטנה יותר  $\Leftrightarrow$  רווח סמך יקטן.
- כאשר  $n < 30$  ניתן להניח נורמליות.
- אם  $L \leq \theta \leq U$  ברמת סמך  $1-\alpha$  אזי רמת הסמך ל- $g(\theta)$  ברמת סמך  $1-\alpha$  הוא  $g(L) \leq g(\theta) \leq g(U)$ .
- קירוב של בינומי לנורמלי:  $X \sim \text{bin}(n, p)$  אזי:  $x \sim N(np, np(1-p))$
- כאשר המדגם הוא מדגם מזווג (2 אוכלוסיות כאשר האחת תלויה בשנייה) נגדיר:  $D_i = X_i - Y_i$ , האמד יהיה אמד לתוחלת של  $D_i$ . (בטבלה – paired observations).

**אמידה בשיטת הריבועים הפחותים:**

מחפשים את  $b_0$  (אמד ל- $\beta_0$ ) ואת  $b_1$  (אמד ל- $\beta_1$ ) כך שסכום הריבועיות של  $y$  מ-

$$\min \sum (y_i - [b_0 + b_1 x_i])^2$$

הוא מינימלי, כלומר:  $e_i$  הוא הסטייה של  $y_i$  מהקו שאמדנו. לעומת זאת  $\varepsilon_i$  הוא הסטייה של  $y_i$  מקו התוחלות האמיתי שאיננו יודעים אותו.

**אמדי ריבועים פחותים:**

$$\hat{\beta}_1 = b_1 = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = b_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = \frac{1}{n-2} [S_{yy} - \hat{\beta}_1^2 S_{xx}]$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y}$$

$$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$$

$$S_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2$$

קו הריבועים הפחותים:  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x$  (תמיד יעבור בנקודה  $(\bar{x}, \bar{y})$ ).

**בדיקת השערות:**

השערת האפס	סטטיסטי המבחן	השערה אלטרנטיבית	אזור דחייה
$H_0: \beta_1 > 0$	$T = \frac{\hat{\beta}_1 - 0}{\hat{\sigma} / \sqrt{S_{xx}}}$	$H_1: \beta_1 > 0$	$T > t_{1-\alpha}^{(n-2)}$
		$H_1: \beta_1 < 0$	$T < t_{1-\alpha}^{(n-2)}$
		$H_1: \beta_1 \neq 0$	$ T  > t_{1-\frac{\alpha}{2}}^{(n-2)}$

ההשערה  $H_0: \beta_1 = 0$ , לעומת  $H_1: \beta_1 \neq 0$  בודקת את מבוהקות הרגרסיה

כאשר השונות של התצפיות שונות נשתמש בשיטת הנראות המקסימלית ולא בשיטת קו הריבועים הפחותים:

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

$$L(\beta_0, \beta_1) = \prod \left( \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2}} \right)$$

הנראות על מנת להקל על החישוב. לאחר מכן נגזור ע"פ  $\beta_1$  ו- $\beta_0$ , נשווה לאפס ונחלץ את המשתנים. נגזור שנית על מנת לוודא מקסימום.

$$\hat{\sigma}^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{n} \leq \hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)$$

**(המשך):**

**אמידת תוחלת Y עבור X0 ורווח סמך  $\alpha$ :**

$$Y_0 \in [\hat{Y}_0 \pm \sigma \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{xx}}} \cdot Z_{1-\frac{\alpha}{2}}]$$

$$Y_0 \in [\hat{Y}_0 \pm \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{xx}}} \cdot t_{1-\frac{\alpha}{2}}^{(n-2)}]$$

**הסקה (בהנחת נורמליות  $Y_i$ ):**

$$\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{S_{xx}}) \Rightarrow \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}} \sqrt{S_{xx}} \sim t(n-2)$$

$$\hat{\beta}_0 \sim N(\beta_0, \frac{\sigma^2 \sum X_i^2}{n \cdot S_{xx}}) \Rightarrow \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\frac{\hat{\sigma}^2 \sum x_i^2}{n S_{xx}}}} \sim t(n-2)$$

מכאן ניתן לחשב אמדים לשוניות של האמדים.

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} \sim \frac{\sigma^2}{n-2} \chi^2(n-2) \Rightarrow \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$$

**רווח סמך ל- $\beta$  ברמת סמך  $1-\alpha$ :**

$$\hat{\beta}_1 - t_{1-\frac{\alpha}{2}}^{(n-2)} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + t_{1-\frac{\alpha}{2}}^{(n-2)} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

מקדם מתאם לינארי:  $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \hat{\beta}_1 \sqrt{\frac{S_{xx}}{S_{yy}}}$

פרופורצית השונות המוסברת (כמה אחוז מהשונות במשתנה המוסבר Y מוסברת ע"י המשתנה המסביר X)  $r^2$

**רווח תחזית:**

$$Y_0 \in [\hat{Y}_0 \pm \sigma \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{xx}}} \cdot Z_{1-\alpha}]$$

$$Y_0 \in [\hat{Y}_0 \pm \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{xx}}} \cdot t_{1-\frac{\alpha}{2}}^{(n-2)}]$$

רווח הסמך מדבר על הנתונים שבגבולות המדגם בלבד. לעומת זאת רווח התחזית לא בהכרח מדבר על הטווח של המדגם.