מ"מ רציפים

	פונקצית צפיפות	פונקצית התפלגות/ תוספות	$E(x)$	$E(x^2)$	$Var(x)$
התפלגות אחידה $x\sim U([L,R])$	$f_X(x)=\begin{cases}0 & x<L\\ \frac{1}{R-L} & L\le x\le R\\ 0 & R<x\end{cases}$	$F_X(x)=\begin{cases}0 & x<L\\ \frac{x-L}{R-L} & L\le x\le R\\ 1 & R<x\end{cases}$	$\frac{R+L}{2}$		$\frac{(R-L)^2}{12}$
התפלגות מעריכית $x\sim exp(\lambda)$ $\lambda>0$	$f_X(x)=\begin{cases}\lambda e^{-\lambda x} & x\ge 0\\ 0 & o.w\end{cases}$	$F_X(x)=\begin{cases}1-e^{-\lambda x} & x\ge 0\\ 0 & o.w\end{cases}$	$\frac{1}{\lambda}$	$\frac{2}{\lambda^2}$	$\frac{1}{\lambda^2}$
התפלגות גאמה $x\sim Gamma(\alpha,\lambda)$ $\alpha>0\ ,\lambda>0$	$f_X(x,\alpha,\lambda)=\begin{cases}\frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} & x\ge 0\\ 0 & o.w\end{cases}$	$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx$ $\Gamma(1)=1,\quad \Gamma(0.5)=\sqrt{\pi}$ $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$ (רק אם α שלמה) $\Gamma(\alpha)=(\alpha-1)!$	$\frac{\alpha}{\lambda}$	$\frac{\alpha(1+\alpha)}{\lambda^2}$	$\frac{\alpha}{\lambda^2}$
התפלגות נורמאלית סטנדרטית $Z\sim N(0,1)$	$f_Z(z,\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$	**<u>טרנספורמציה</u>** $x=\sigma Z+\mu$ $\frac{x-\mu}{\sigma}\sim N(0,1)$ $\overline{x_n}\sim N(\mu,\frac{\sigma^2}{n})$ $\frac{\overline{x_n}-\mu}{\frac{\sigma^2}{n}}\sim N(0,1)$ נכון לכל ממוצע מכל התפלגות כש- n גדול לפי משפט הגבול המרכזי.	0	1	1
התפלגות נורמאלית $x\sim N(\mu,\sigma^2)$ $-\infty<\mu<\infty$ $\sigma^2>0$	$f_X(x,\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	**<u>סכום משתנים נורמאליים:</u>** $x\sim N(\mu_1,{\sigma_1}^2)\to\tilde{x}=\frac{x-\mu_1}{\sigma_1}\to\tilde{x}\sim N(0,1)$ $y\sim N(\mu_2,{\sigma_2}^2)\to\tilde{y}=\frac{y-\mu_2}{\sigma_2}\to\tilde{y}\sim N(0,1)$ $S=x+y\sim N(\mu_1+\mu_2,{\sigma_1}^2+{\sigma_2}^2)$ $\tilde{S}=\tilde{x}+\tilde{y}\sim N(0,2)$	μ	$\sigma^2+\mu^2$	σ^2
התפלגות לוג נורמאלית $y\sim LogNormal(\mu,\sigma^2)$ $x=log_e y\sim N(\mu,\sigma^2)$ $y=e^x$	$f_Y(y)=\begin{cases}\frac{1}{y\sqrt{2\pi\sigma^2}}e^{-\frac{(log_e y-\mu)^2}{2\sigma^2}} & x\ge 0\\ 0 & o.w\end{cases}$	**<u>תקנון לנורמאלי סטנדרטי:</u>** $P(y\le m)=P(e^x\le m)=$ $=P(x\le log_e m)=P(Z\le\frac{log_e m-\mu}{\sigma})$	$e^{\mu+\frac{1}{2}\sigma^2}$		$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$
התפלגות $chi\ squared$ $Z\sim N(0,1)$ $Y=Z^2\sim{\chi^2}_{(1)}=gamma(\frac{1}{2}\frac{1}{2})$	$f_Y(y)=\begin{cases}\frac{\frac{1}{2}^{\frac{1}{2}}y^{-\frac{1}{2}}e^{-\frac{1}{2}x}}{\Gamma(\frac{1}{2})} & x\ge 0\\ 0 & o.w\end{cases}$	$\Gamma(0.5)=\sqrt{\pi}$	1		2
סכום התפלגות $chi\ squared$ $Z_i\sim N(0,1)$ $\sum_{i=1}^n {Z_i}^2\sim{\chi^2}_{(n)}=gamma(\frac{n}{2}\frac{1}{2})$	$x_i\sim N(\mu,\sigma^2)$ $\sum_{i=1}^n\left(\frac{x_i-\mu}{\sigma^2}\right)^2=\sum_{i=1}^n {Z_i}^2\sim{\chi^2}_{(n)}$	$\sum_{i=1}^n\left(\frac{x_i-\overline{x}_n}{\sigma^2}\right)^2=\sum_{i=1}^n {Z_i}^2\sim{\chi^2}_{(n-1)}$	n		$2n$

	פונקצית הצפיפות	פונקצית הצפיפות המצטברת	$E(x)$	$E(x^2)$	$Var(x)$
כללי	$P(a < x \leq b) = \int_a^b f_X(x)dx$	$F_X(x) = \int_{-\infty}^{x} f_T(t)dt = P(X \leq x)$ **הערה:** כאשר בונים את הקטע הרלוונטי, לא לשכוח את החלק שקדם לו שכן מדובר בפונקציה מצטברת.	$\int_{-\infty}^{\infty} x f_X(x)dx$	$\int_{-\infty}^{\infty} x^2 f_X(x)dx$	$\int_{-\infty}^{\infty} x^2 f_X(x)dx - [E(x)]^2$
תכונות	.1 $f_X(x) \geq 0$ לכל x .2 $\int_{-\infty}^{\infty} f_X(x)dx = 1$	.1 $F_X(x)$ מונוטונית לא יורדת $F_X(a) > F_X(b) \leftarrow a > b$ .2 $\text{Lim}_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to\infty} F_X(x) = 1$ .3 $F_X(x)$ רציפה מימין: $\lim_{x\to a+} F_X(x) = F_X(a)$			

דברים נוספים

<u>דוגמה מ- מ"מ שאינו מתפלג אחיד $[0,1]$</u>	<u>קונבולוציה:</u> $X_1 + X_2 = Y$	<u>מינימום ומקסימום של מ"מ</u>	<u>אי שוויון צ'בישב</u>
$x \sim F$ ו- $U \sim U([0,1])$ נגדיר: $F_x(x) = U$ (פונקצית ההתפלגות) נחלץ את x: $x = F_x^{-1}(U)$ נגריל מס' בין 0 ל- 1, נציב ב- U ונתרגם להתפלגות F	$F_Y(y) = \int_{-\infty}^{y} F_{x_1}(u) \times F_{x_2}(y-u)\, du$	$max(x_i) = [F_x(x)]^n$ $min(x_i) = 1 - [1 - F_x(x)]^n$	$p(\lvert x - E(x)\rvert > a) < \frac{v(x)}{a^2}$

מ"מ בדידים

	התפלגות	סימון	תוחלת	שונות	פונקצית ההסתברות	מתי נשתמש?
	אחידה	$U(M,N)$	$\frac{M+N}{2}$	$\frac{(N-M+1)^2-1}{12}$	$P(X=k) = \frac{1}{N-M+1},\quad k = M, M+1, \dots N-1, N$	משתנה בו כל ערכי המשתנה שווי הסתברות.
בלתי תלויים + עם החזרה	ברנולי	$Ber(p)$	p	$p(1-p)$	$P(X=1) = p$ $P(X=0) = 1-p$	משתנה בינארי שבו יש ניסוי יחיד. אם תוצאת היעד התרחשה, המשתנה מקבל את הערך 1, או 0 אחרת.
	בינומי	$Bin(n,p)$	np	$np(1-p)$	$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}\quad k = 0,1,\dots,n$	משתנה המחשב הסתברות של k הצלחות ב- n ניסויי ברנולי ב"ת ושווי התפלגות.
	גיאומטרי	$Geo(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$P(X=k) = p(1-p)^{k-1},\ k = 1,2,\dots,\infty$	משתנה הסופר מס' ניסויי ברנולי ב"ת ושווי התפלגות עד ההצלחה הראשונה.
	בינומי שלילי	$NB(n,p)$	$\frac{n}{p}$	$\frac{n(1-p)}{p^2}$	$P(X=k) = \binom{k-1}{n-1} p^n (1-p)^{k-n},$ $k = n, n+1, n+2, \dots, \infty$	משתנה הסופר מס' ניסויים בינומיים ב"ת ושווי התפלגות עד ההצלחה ה- n (ולא עד ההצלחה הראשונה כמו במ"מ גיאומטרי).
ללא החזרה + ללא סדר	היפר גיאומטרי	$HG(n,A,B)$	$\frac{nA}{A+B}$	$\frac{nAB(A+B-n)}{(A+B)^2(A+B-1)}$	$P(X=k) = \frac{\binom{A}{k}\binom{B}{n-k}}{\binom{A+B}{n}},$ B	משתנה הסופר מדגם בגודל n מתוך אוכלוסיה בת N איברים זמינים. נשמש במשתנה זה כאשר נתונה לנו אוכלוסיה בגודל N שמתחלקת ל- 2 קטגוריות: A ו- B וכן נתון לנו מה המדגם.
	פואסון	$Pois(\lambda)$	λ	λ	$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!},\quad k \geq 0$	משתנה הסופר את מידת הצפיפות ליחידה סטנדרטית. שאלות שהנתון בהן הוא ממוצע האירועים ליחידת זמן או שטח.

אמידה- הקצאה אופטימאלית

מה הסיפור?	מה עושים?	נוסחאות רלוונטיות	הערות
ברצוננו למצוא אומד לפרמטר לא ידוע. עלינו לבחור לביצוע אחד משני סוגי המדגמים המוצעים.	בודקים למי מהשניים MSE נמוך יותר ויוצרים ממנו מדגם בעל n תצפיות.	$\boldsymbol{MSE} = E\left[\left(\hat{\theta}-\theta\right)^2\right]$ $= \boldsymbol{Var}(\hat{\boldsymbol{\theta}}) + \boldsymbol{bias}^2(\hat{\boldsymbol{\theta}})$ $= E\left[\left(\hat{\theta}-E(\theta)\right)\right]^2 + [E(\hat{\theta})-\theta]^2$	1. אומד ייקרא חסר הטיה אם: $E(\hat{\theta}) = \theta$. 2. אומד ייקרא עקיב אם: $MSE(\hat{\theta}_n)_{n\to\infty}\xrightarrow{}0$. 3. אומד ייקרא אומד מוטה אסימפטוטית אם ככל ש- n הולך וגדל- ההטיה קטנה.
נתונים שני מדגמים שכבר התקיימו, על מנת למצוא אומד לפרמטר לא ידוע. ברצוננו לקבוע איזה משקל לתת לכל מדגם באומד המשוקלל.	האומד יורכב כך: $\hat{\mu} = \alpha\overline{x_n} + (1-\alpha)\overline{y_n}$ כיוון שהאומד הוא חסר הטיה, נמצא את ה- α שממזערת את השונות על ידי גזירה.	האומד: $\hat{\mu} = \alpha\overline{x_n} + (1-\alpha)\overline{y_n}$ שונות האומד: $Var(\hat{\mu}) = \alpha^2\frac{\sigma_1^2}{n_1} + (1-\alpha)^2\frac{\sigma_2^2}{n_2}$ לאחר גזירה: $\alpha = \frac{\frac{\sigma_2^2}{n_2}}{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$	
ברצוננו לאמוד פרמטר לא ידוע. הפעם אנו חייבים להשתמש בשני מדגמים כיוון שהפרמטר הרצוי מורכב משני פרמטרים- אחד מכל מדגם, או שכל מדגם מורכב משני פרמטרים לא ידועים, מה שמחייב שני מדגמים כדי "לקזז" את הנעלם הלא רצוי. כאן השאלה היא של מחירי הדגימה וחלוקת התקציב בין שני המדגמים.	אם ברשותנו תקציב ל- n דגימות נקבע שרירותית של- x נקצה m דגימות ול- y נקצה $n-m$. תוך שימוש בממוצעי המדגם ניצור אומד חסר הטיה, ואת שונות האומד נגזור לפי m. לבסוף נגיע למשוואה ריבועית של m, אשר פתרונה ימצא לנו את אחוז התצפיות של m שיש להקצות ל- x מתוך n.		אם נתון שמחיר כל דגימה מ- x כפול משל y, עבור x נקצה m תצפיות ועבור y נקצה $n-2m$.

חוקי תוחלות, שונויות ולוגריתמים

	תוחלת $E(x)$	שונות $Var(x)$ – מדגמים ב"ת	שונות $Var(x)$ – מדגמים תלויים	שונות משותפת $cov(x,y)$
סכום	תוחלת של סכום שווה לסכום התוחלות: $E(x+y) = E(x) + E(Y)$	שונות של סכום שווה לסכום השונויות: $Var(x+y) = Var(x) + Var(y)$	$Var(x+y) = Var(x) + Var(y) + 2cov(x,y)$	$cov(x+z,y) = cov(x,y) + cov(z,y)$
		דגש: $Var(x-y) = Var(x) + Var(-y) = Var(x) + (-1)^2 var(Y) = var(x) + var(Y)$		
הכפלה בקבוע (ליניאריות)	$E(ax) = aE(x)$	$Var(ax) = a^2 Var(x)$		$cov(ax,by) = a\times b\times cov(x,y)$
קבוע	$E(a) = a$	$Var(a) = 0$		$cov(a,b) = 0$
כתוצאה מכך	$E(\overline{x_n}) = \frac{n\mu}{n} = \mu$	$Var(\overline{x_n}) = \frac{\sigma^2}{n}$		

לוגריתמים
$\log(a\times b) = \log(a) + \log(b)$
$\log\left(\frac{a}{b}\right) = \log(a) - \log(b)$
$\log(a^b) = b\times\log(a)$
$\log(e^a) = a$
$[\log(x)]' = \frac{1}{x}$

	x	$\overline{x_n}$
$E(x)$	$E(x) = \mu$	$E(\overline{x_n}) = \mu$
$v(x)$	$v(x) = \sigma^2$	$V(\overline{x_n}) = \frac{\sigma^2}{n}$
$E(x^2)$	$E(x^2) = \mu^2 + \sigma^2$	$E(\overline{x_n}) = \mu^2 + \frac{\sigma^2}{n}$

שיטות לאמידה

שיטת המומנטים	$$\mu_k = \hat{E}(x^k) = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$	ניקח את המומנט הראשון שהוא פונקציה של הפרמטר הנאמד.
שיטת הנראות המקסימאלית	**פונקצית הנראות:** $$L(x,\theta) = P(X_i = x_i) = \prod_{i=1}^{n}(\text{פונקציית הצפיפות/ההסתברות בהתאם להתפלגות})$$ כדי להיפטר מהמכפלות נעשה **טרנספורמציה של log_e** על פונקצית הנראות, כיוון שלפי חוקי הלוגים, כפל בתוך לוג שקול לסכימה של לוגים. כך נעבור מ- $$\log_e L(x,\theta) = log_e \prod_{i=1}^{n}(\text{פונקציית הצפיפות/ההסתברות בהתאם להתפלגות})$$ ל- $= \sum_{i=1}^{n} log_e(\text{פונקציית הצפיפות/ההסתברות בהתאם להתפלגות})$ נפשט את הביטוי, **ונגזור לפי θ עד למציאת האומד.**	אומד בשיטה זו הוא זה שמביא למקסימום את פונקצית הנראות
	לעיתים לא נוכל לעשות טרנספורמציה של לוג על פונקצית הנראות ולגזור כדי להגיע לתשובה. אז יש לחשוד ש- $\hat{\theta} = minx_i/maxx_i$ או וריאציה דומה, כלומר שהמקסימום של פונקצית הנראות נמצא בקצוות הקטע. כלל אצבע לכך יהיה כאשר מדובר בפונקציה עם תחום מפוצל, כאשר התומך תלוי ב- θ. במצב זה נחליף את התומך למשל מ- $\prod_{i=1}^{n}(x_i \le \theta)$ ל- $maxx_i \le \theta$ (אם כל ה- x-ים צריכים להיות קטנים מ- θ, מספיק לדרוש שה- x המקסימאלי יהיה קטן מ- θ). כעת נסתכל על פונקצית הנראות עצמה. אם מדובר בפונקציה יורדת של θ, המקסימום של פונקצית הנראות מתקבל ככל ש- θ קטנה יותר. במקרה זה ה- θ הכי קטנה שנוכל להתיר לעצמינו היא זו שמקיימת את דרישת האינדיקטור $maxx_i \le \theta$, ואז האומד יהיה $maxx_i = \theta$. (בהתאמה ניתן לעשות גם עם מינימום). דגש : פונקצית הנראות היא פונקציה של θ. ה- x-ים הם קבועים. (כדי למצוא MSE נמצא את פונקצית הצפיפות המצטברת של האומד לפי הנוסחאות למינ׳/ מקס׳ של מ״מ רציף, ועל ידי גזירה נגיע לפונקצית הצפיפות. אז נוכל לחשב את ההטיה והשונות).	

רווח סמך

		התפלגות	רווח הסמך	הערות
רווח סמך לתוחלת מקור התצפיות הן מהתפלגות נורמאלית $x_i \sim N(\mu,\sigma^2)$ **או n גדול מאוד (משפט הגבול המרכזי)**	**שונות ידועה**	$\frac{\bar{x}_n-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$	$P\left(\bar{x}_n - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} < \mu < \bar{x}_n + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right) = 1-\alpha$	1. **גודל האוכלוסייה**- ככל שיותר גדולה, השונות יותר קטנה ור"ס קטן. 2. **גודל המדגם**- ככל שיותר גדול ר"ס קטן. 3. **רמת הביטחון $(1-\alpha)$**- ככל שיותר גדולה כך ר"ס קטן. 4. **אורך רווח הסמך** אינו תלוי ב- $\bar{x}_n$ $L_{\text{ר"ס}} = 2 \times Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}$
	שונות לא ידועה	$\frac{\bar{x}_n-\mu}{\sqrt{\frac{{s_n}^2}{n}}} \sim T_{(n-1)}$	$P\left(\bar{x}_n - T_{1-n}^{1-\frac{\alpha}{2}}\sqrt{\frac{{s_n}^2}{n}} < \mu < \bar{x}_n + T_{n-1}^{1-\frac{\alpha}{2}}\sqrt{\frac{{s_n}^2}{n}}\right) = 1-\alpha$	
רווח סמך להפרש תוחלות $x_i \sim N(\mu_x,{\sigma_x}^2)$ $y_i \sim N(\mu_y,{\sigma_y}^2)$	**נתונים מזווגים במדגמים תלויים** – **שונות ידועה/לא ידועה**	$\frac{\bar{d}_n-\mu_d}{\sqrt{\frac{{s_d}^2}{n}}} \sim T_{(n-1)}$	$d_i = x_i - y_i$ $P\left(\bar{d}_n - T_{1-n}^{1-\frac{\alpha}{2}}\sqrt{\frac{{s_d}^2}{n}} < \mu_d < \bar{d}_n + T_{n-1}^{1-\frac{\alpha}{2}}\sqrt{\frac{{s_d}^2}{n}}\right) = 1-\alpha$	
	נתונים לא מזווגים במדגמים ב"ת – **שונויות ידועות ושוות**	$\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma^2}{n_x}+\frac{\sigma^2}{n_y}}} \sim N(0,1)$	$P\left((\bar{x}-\bar{y}) - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n_x}+\frac{\sigma^2}{n_y}} < \mu_x-\mu_y < (\bar{x}-\bar{y}) + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n_x}+\frac{\sigma^2}{n_y}}\right) = 1-\alpha$	
	נתונים לא מזווגים במדגמים ב"ת – **שונויות לא ידועות אך שוות**	$\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{{s_p}^2}{n_x}+\frac{{s_p}^2}{n_y}}} \sim T_{(n-1)}$	$P\left((\bar{x}-\bar{y}) - T_{1-n}^{1-\frac{\alpha}{2}}\sqrt{\frac{{s_p}^2}{n_x}+\frac{{s_p}^2}{n_y}} < \mu_x-\mu_y < (\bar{x}-\bar{y}) + T_{1-n}^{1-\frac{\alpha}{2}}\sqrt{\frac{{s_p}^2}{n_x}+\frac{{s_p}^2}{n_y}}\right) = 1-\alpha$	
רווח סמך ל- P $x_i \sim Ber(p)$	**פרופורציה**	$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$	$P\left(\hat{p} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1-\alpha$	1. $\hat{p}$- הפרופורציה במדגם (ממוצע המדגם). 2. $\hat{p}(1-\hat{p})$- שונות של ברנולי 3. אם n גדול מאוד אז $\hat{p} \approx \frac{1}{2}$
	הפרש פרופורציות		$P\left((\widehat{p_1}-\widehat{p_2}) - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1}+\frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}} < p_1-p_2 < (\widehat{p_1}-\widehat{p_2}) + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1}+\frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}\right) = 1-\alpha$	
	Wilson		$P\left(\frac{\hat{p}+\frac{Z^2_{1-\frac{\alpha}{2}}}{2n} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{Z^2_{1-\frac{\alpha}{2}}}{4n^2}}}{1+\frac{Z^2_{1-\frac{\alpha}{2}}}{n}} < p < \frac{\hat{p}+\frac{Z^2_{1-\frac{\alpha}{2}}}{2n} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{Z^2_{1-\frac{\alpha}{2}}}{4n^2}}}{1+\frac{Z^2_{1-\frac{\alpha}{2}}}{n}}\right) = 1-\alpha$	
רווח סמך לשונות $x_i \sim N(\mu,\sigma^2)$	**התוחלת ידועה**	$\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2} \sim {\chi^2}_{(n)}$	$P\left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{{\chi^2}_{(n),\,1-\frac{\alpha}{2}}} < \sigma^2 < \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{{\chi^2}_{(n),\,\frac{\alpha}{2}}}\right) = P\left(\frac{n\times\hat{\sigma}^2}{{\chi^2}_{(n),\,1-\frac{\alpha}{2}}} < \sigma^2 < \frac{n\times\hat{\sigma}^2}{{\chi^2}_{(n),\,\frac{\alpha}{2}}}\right) = 1-\alpha$	
	התוחלת אינה ידועה	$\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2} \sim {\chi^2}_{(n-1)}$	$P\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{{\chi^2}_{(n-1),\,1-\frac{\alpha}{2}}} < \sigma^2 < \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{{\chi^2}_{(n-1),\,\frac{\alpha}{2}}}\right) = P\left(\frac{(n-1)\times s^2}{{\chi^2}_{(n-1),\,1-\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)\times s^2}{{\chi^2}_{(n-1),\,\frac{\alpha}{2}}}\right) = 1-\alpha$	

	אומד לשונות (לפי נראות מירבית ושיטת המומנטים)- אומד מוטה	אומד לשונות- אומד בלתי מוטה	אומד לשונות המתואמת
אומד לשונות $x_i \sim N(\mu,\sigma^2)$	$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i)^2 - \frac{n}{n-1}{\bar{x}_n}^2 = \frac{n}{n-1}\hat{\sigma}^2$	$S^2_{pooled} = \frac{(n_x-1)s_x^2+(n_y-1)s_y^2}{n_x+n_y-2}$ $= \frac{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2+\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}{n_x+n_y-2}$
	$\frac{n\times\hat{\sigma}^2}{\sigma^2} \sim {\chi^2}_{(n)} = gamma\left(\frac{n}{2},\frac{1}{2}\right)$	$\frac{(n-1)\times s^2}{\sigma^2} \sim {\chi^2}_{(n-1)} = gamma\left(\frac{n-1}{2},\frac{1}{2}\right)$	

בדיקת השערות

מה בודקים?			חד צדדי	דו צדדי	הערות
בדיקת השערות בדבר תוחלת של מ"מ נורמאלי או כמות תצפיות גדולה במיוחד. $x\sim N(\mu,\sigma^2)$		מבנה המבחן	$H_0:\mu=\mu_0$ $H_1:\mu\gtrless\mu_0$	$H_0:\mu=\mu_0$ $H_1:\mu\neq\mu_0$	
	שונות ידועה	אזור הדחיה	$R=\left\{\overline{X_n}\gtrless\mu_0\pm Z_{1-\alpha}\sqrt{\frac{\sigma^2}{n}}\right\}$	$R=\begin{cases}\overline{X_n}>\mu_0+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\\ \overline{X_n}<\mu_0-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\end{cases}$	
	שונות לא ידועה	אומד בלתי מוטה לשונות	$\hat\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar x_n)^2=\frac{1}{n}\sum_{i=1}^{n}(x_i)^2-\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2 \rightarrow s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar x_n)^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i)^2-\frac{n}{n-1}\bar x_n^2=\frac{n}{n-1}\hat\sigma^2$		
		אזור הדחיה	$R=\left\{\overline{X_n}\gtrless\mu_0+T_{n-1}^{1-\alpha}\sqrt{\frac{s^2}{n}}\right\}$	$R=\begin{cases}\overline{X_n}>\mu_0+T_{n-1}^{1-\frac{\alpha}{2}}\sqrt{\frac{s^2}{n}}\\ \overline{X_n}<\mu_0-T_{n-1}^{1-\frac{\alpha}{2}}\sqrt{\frac{s^2}{n}}\end{cases}$	
בדיקת השערות בדבר שוויון תוחלות של מ"מ נורמאליים או כמות תצפיות גדולה במיוחד. $x\sim N(\mu_x,\sigma^2)$ $y\sim N(\mu_y,\sigma^2)$	מדגמים לא מזווגים ובלתי תלויים	מבנה המבחן	$H_0:\delta=0$ $H_1:\delta\gtrless 0$ $H_0:\mu_x=\mu_y$ $H_1:\mu_x\gtrless\mu_y$	$H_0:\delta=0$ $H_1:\delta\neq 0$ $H_0:\mu_x=\mu_y$ $H_1:\mu_1\neq\mu_2$	$\mu_x-\mu_y=\delta$
	שונות ידועות ושוות	אזור הדחיה	$R=\left\{\hat\delta\gtrless 0\pm Z_{1-\alpha}\sqrt{\frac{\sigma^2}{n_x}+\frac{\sigma^2}{n_y}}\right\}$	$R=\begin{cases}\hat\delta>0+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n_x}+\frac{\sigma^2}{n_y}}\\ \hat\delta<0-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n_x}+\frac{\sigma^2}{n_y}}\end{cases}$	$\overline{X_{n_x}}-\overline{Y_{n_y}}=\hat\delta$
	שונות לא ידועות אך שוות	שונות מתואמת	$S^2_{pooled}=\frac{(n_x-1)s_x^2+(n_y-1)s_y^2}{n_x+n_y-2}=\frac{\sum_{i=1}^{n}(x_i-\bar x_n)^2+\sum_{i=1}^{n}(y_i-\bar y_n)^2}{n_x+n_y-2}$		$s_x^2=\frac{1}{n_x-1}\sum_{i=1}^{n}(x_i)^2-\frac{n_x}{n_x-1}\bar x_n^2$
		אזור הדחיה	$R=\left\{\hat\delta\gtrless 0\pm T_{n_x+n_y-2}^{1-\alpha}\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}}\right\}$	$R=\begin{cases}\hat\delta>0+T_{n_x+n_y-2}^{1-\frac{\alpha}{2}}\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}}\\ \hat\delta<0-T_{n_x+n_y-2}^{1-\frac{\alpha}{2}}\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}}\end{cases}$	
	מדגמים מזווגים ותלויים $d_i=y_i-x_i$	מבנה המבחן	$H_0:\mu_d=0$ $H_1:\mu_d\gtrless 0$ $H_0:\mu_x=\mu_y$ $H_1:\mu_x\gtrless\mu_y$	$H_0:\mu_d=0$ $H_1:\mu_d\neq 0$ $H_0:\mu_x=\mu_y$ $H_1:\mu_x\neq\mu_y$	$\mu_x-\mu_y=\mu_d$
		אומד בלתי מוטה לשונות	$\hat\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(d_i-\bar d_n)^2=\frac{1}{n}\sum_{i=1}^{n}(d_i)^2-\left(\frac{1}{n}\sum_{i=1}^{n}d_i\right)^2 \rightarrow s^2=\frac{1}{n-1}\sum_{i=1}^{n}(d_i-\bar d_n)^2=\frac{1}{n-1}\sum_{i=1}^{n}(d_i)^2-\frac{n}{n-1}\bar d_n^2=\frac{n}{n-1}\hat\sigma^2$		
		אזור הדחיה	$R=\left\{\overline{d_n}\gtrless 0\pm T_{n-1}^{1-\alpha}\sqrt{\frac{s_d^2}{n}}\right\}$	$R=\begin{cases}\overline{d_n}>0+T_{n-1}^{1-\frac{\alpha}{2}}\sqrt{\frac{s_d^2}{n}}\\ \overline{d_n}<0-T_{n-1}^{1-\frac{\alpha}{2}}\sqrt{\frac{s_d^2}{n}}\end{cases}$	$\overline{X_{n_x}}-\overline{Y_{n_y}}=\overline{d_n}$ אם אין הנחת נורמאליות במדגמים המזווגים, נבדוק את ההשערה בעזרת מבחן הסימן.

<table>
<tr><th colspan="3">מה בודקים?</th><th colspan="2">חד צדדי</th><th colspan="2">דו צדדי</th><th>הערות</th></tr>
<tr><td rowspan="7">בדיקת השערות בדבר <u>פרופורציה</u> של מ"מ בינומי
$x\sim Bin(n,p)$
(הדבר גם נכון לכל התפלגות בדידה אחרת)</td><td colspan="2">מבנה המבחן</td><td colspan="2">$H_0: p=p_0$
$H_1: p\gtrless p_0$</td><td colspan="2">$H_0: p=p_0$
$H_1: p\neq p_0$</td><td></td></tr>
<tr><td rowspan="2">קירוב נורמאלי- כמות תצפיות גדולה במיוחד
$n\geq 30$</td><td rowspan="2">אזור הדחיה</td><td colspan="2">$R=\left\{\overline{X_n}\gtrless np_0\pm Z_{1-\alpha}\sqrt{np_0(1-p_0)}\right\}$</td><td colspan="2">$R=\begin{cases}\overline{X_n}>np_0+Z_{1-\frac{\alpha}{2}}\sqrt{np_0(1-p_0)}\\ \overline{X_n}<np_0-Z_{1-\frac{\alpha}{2}}\sqrt{np_0(1-p_0)}\end{cases}$</td><td rowspan="2"></td></tr>
<tr><td colspan="2">$R=\left\{\hat{p}\gtrless p_0\pm Z_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n}}\right\}$</td><td colspan="2">$R=\begin{cases}\hat{p}>p_0+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_0(1-p_0)}{n}}\\ \hat{p}<p_0-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_0(1-p_0)}{n}}\end{cases}$</td></tr>
<tr><td rowspan="4">מספר התצפיות אינו מספק כדי להשתמש בקירוב הנורמאלי של משפט הגבול המרכזי</td><td rowspan="2">מבנה המבחן</td><td><u>חד צדדי ימני</u></td><td><u>חד צדדי שמאלי</u></td><td><u>דו צדדי סימטרי</u></td><td><u>דו צדדי לא סימטרי</u></td><td rowspan="4">1. במבחן חד צדדי נחפש את סכום ההסתברויות שלא עולה על α
2. במבחן דו צדדי נחפש בכל אחד משני הצדדים של ההתפלגות את סכום ההסתברויות שלא עולה $\frac{\alpha}{2}$ ויחד נגיע לרמת מובהקות של α .
3. במבחן דו צדדי כאשר $p=\frac{1}{2}$ תחת H_0, תהיה סימטריה בהתפלגות ובין שני הערכים הקריטיים.
4. במקרה הבדיד אי אפשר לבחור רמת מובהקות ספציפית, כי ההסתברויות אינן רציפות . אם הדרישה היא למקסימום $\alpha=0.05$ ייתכן שנאלץ להתפשר.
5. כשמחפשים את ה- $p-value$ נחשב את ההסתברות לקבל את התוצאה שקיבלנו בניסוי+ ההסתברות לקבל תוצאות קיצוניות יותר לכיוון H_1.</td></tr>
<tr><td>$H_0: p=p_0$
$H_1: p>p_0$</td><td>$H_0: p=p_0$
$H_1: p<p_0$</td><td>$H_0: p=\frac{1}{2}$
$H_1: p\neq\frac{1}{2}$</td><td>$H_0: p=p_0$
$H_1: p\neq p_0$</td></tr>
<tr><td>אזור הדחיה</td><td>$R=\{x>c_1\}$</td><td>$R=\{x<c_2\}$</td><td colspan="2">$R=\begin{cases}x>c_1\\ x<c_2\end{cases}$</td></tr>
<tr><td>איך מוצאים את הערך הקריטי?</td><td colspan="4"><table>
<tr><th>c</th><th>0</th><th>1</th><th>2</th><th>.......</th><th>28</th><th>29</th></tr>
<tr><td>$P_{p_0}(x=c)$</td><td colspan="6">הסתברות התוצאות לפי פונקציית ההסתברות של מ"מ בינומי</td></tr>
<tr><td>$P_{p_0}(x<c_2)\leq\alpha$
(חד צדדי שמאלי)</td><td colspan="2">נסכום את ההסתברויות עד שנגיע ל- α הרצויה</td><td></td><td></td><td></td><td></td></tr>
<tr><td>$P_{p_0}(x>c_1)\leq\alpha$
(חד צדדי ימני)</td><td></td><td></td><td></td><td></td><td colspan="2">נסכום את ההסתברויות עד שנגיע ל- α הרצויה</td></tr>
<tr><td>$P_{p_0}(x<c_2)+P_{p_0}(x>c_1)\leq\alpha$
(דו צדדי)</td><td colspan="2">נסכום את ההסתברויות עד שנגיע ל- $\frac{\alpha}{2}$ הרצויה</td><td></td><td></td><td colspan="2">נסכום את ההסתברויות עד שנגיע ל- $\frac{\alpha}{2}$ הרצויה</td></tr>
</table></td></tr>
<tr><td rowspan="4">בדיקת השערות בדבר <u>שוויון</u> <u>פרופורציות</u> בין 2 <u>אוכלוסיות</u>:
$x_1\sim Bin(n_1,p_1)$
$x_2\sim Bin(n_2,p_2)$</td><td rowspan="3">מדגמים גדולים</td><td>מבנה המבחן</td><td colspan="2">$H_0: p_1=p_2=p$
$H_1: p_1\gtrless p_2$</td><td colspan="2">$H_0: p_1=p_2=p$
$H_1: p_1\neq p_2$</td><td></td></tr>
<tr><td>אומדים</td><td colspan="2">$\widehat{p_1}=\frac{\text{מס' המיוחדים מאוכלוסיה א'}}{n_1}$</td><td colspan="2">$\widehat{p_2}=\frac{\text{מס' המיוחדים מאוכלוסיה ב'}}{n_2}$</td><td>$\hat{p}=\frac{\text{מס' המיוחדים משתי הדגימות}}{n_1+n_2}$ $\hat{q}=1-\hat{p}$</td></tr>
<tr><td>אזור הדחיה</td><td colspan="2">$R=\left\{\widehat{p_1}-\widehat{p_2}>0\pm Z_{1-\alpha}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{p}\hat{q}}\right\}$</td><td colspan="2">$R=\begin{cases}\widehat{p_1}-\widehat{p_2}>0+Z_{1-\frac{\alpha}{2}}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{p}\hat{q}}\\ \widehat{p_1}-\widehat{p_2}<0-Z_{1-\frac{\alpha}{2}}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{p}\hat{q}}\end{cases}$</td><td>$\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{p}\hat{q}}=\sqrt{\left(\frac{\widehat{p_1}\widehat{q_1}}{n_1}+\frac{\widehat{p_2}\widehat{q_2}}{n_2}\right)}$
הערה : דרישה של רמת מובהקות מקסימאלית גוררת $\hat{p}=\hat{q}=\frac{1}{2}$</td></tr>
<tr><td colspan="2">מדגמים קטנים- Fisher (טבלה 2X2)</td><td colspan="5">פירוט בהרחבה תחת מבחנים א- פרמטריים</td></tr>
</table>

מה בודקים?			חד צדדי	דו צדדי	הערות
בדיקת השערות בדבר <u>שונות</u> של מ"מ נורמאלי או כמות תצפיות גדולה במיוחד. $x\sim N(\mu,\sigma^2)$	מבנה המבחן		$H_0:\sigma^2=\sigma^2{}_0$ $H_1:\sigma^2\gtrless\sigma^2{}_0$	$H_0:\sigma^2=\sigma^2{}_0$ $H_1:\sigma^2\neq\sigma^2{}_0$	
	תוחלת ידועה $\frac{n\times\hat{\sigma}^2}{\sigma^2}\sim\chi^2{}_{(n)}$	**אומד לשונות**	$\hat{\boldsymbol{\sigma}}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$		
		אזור הדחייה	$R=\left\{\frac{n\times\hat{\sigma}^2}{\sigma^2{}_0}\gtrless\chi^2{}_{(n),1-\alpha}\right\}$	$R=\begin{cases}\frac{n\times\hat{\sigma}^2}{\sigma^2{}_0}>\chi^2{}_{(n),z_{1-\frac{\alpha}{2}}}\\ \frac{n\times\hat{\sigma}^2}{\sigma^2{}_0}<\chi^2{}_{(n),z_{1-\frac{\alpha}{2}}}\end{cases}$	
	תוחלת לא ידועה $\frac{(n-1)\times s^2}{\sigma^2}\sim\chi^2{}_{(n-1)}$	**אומד בלתי מוטה לשונות**	$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2=\frac{\mathbf{1}}{\mathbf{n-1}}\sum_{i=1}^{n}(\boldsymbol{x_i})^2-\frac{\boldsymbol{n}}{\boldsymbol{n-1}}\bar{\boldsymbol{x}}_n{}^2=\frac{n}{n-1}\hat{\sigma}^2$		
		אזור הדחייה	$R=\left\{\frac{(n-1)\times s^2}{\sigma^2{}_0}\gtrless\chi^2{}_{(n-1),1-\alpha}\right\}$	$R=\begin{cases}\frac{(n-1)\times s^2}{\sigma^2{}_0}>\chi^2{}_{(n-1),z_{1-\frac{\alpha}{2}}}\\ \frac{(n-1)\times s^2}{\sigma^2{}_0}<\chi^2{}_{(n-1),z_{1-\frac{\alpha}{2}}}\end{cases}$	

<u>טעות מסוג ראשון, טעות מסוג שני, עוצמה ו- $P-Value$</u>

<u>טעות מסוג ראשון</u>	$\boldsymbol{\alpha}=P_{H_0}(H_1)=P_{H_0}(R)$	דחיית H_0 כשהיא למעשה ההשערה הנכונה. זוהי טעות חמורה: קבלת החלטה חדשנית שאין לה הצדקה!
<u>טעות מסוג שני</u>	$\boldsymbol{\beta}=P_{H_1}(H_0)=P_{H_1}(\bar{R})=1-P_{H_1}(R)$	קבלת H_0 כשהיא למעשה ההשערה הלא הנכונה. זוהי טעות פחות חמורה, שכן זוהי למעשה המציאות שהייתה קודם לכן.
<u>העוצמה</u>	$\mathbf{1}-\boldsymbol{\beta}=P_{H_1}(H_1)=P_{H_1}(R)$	דחיית H_0 כאשר היא לא ההשערה הנכונה.

מציאות החלטה	$\boldsymbol{H_0}$	$\boldsymbol{H_1}$
לא לדחות את $\boldsymbol{H_0}$	ההחלטה תואמת מציאות $(1-\alpha)$	טעות מסוג 2 (β)
לדחות את $\boldsymbol{H_0}$	טעות מסוג 1 **<u>רמת המובהקות</u>** $=(\boldsymbol{\alpha})$	ההחלטה תואמת מציאות **<u>העוצמה</u>** $=(\mathbf{1}-\boldsymbol{\beta})$

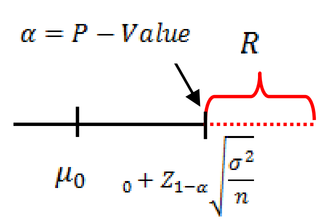
<u>המשמעות של ה- $P-Value$:</u> <u>ההסתברות לקבל את התוצאה שקיבלנו בניסוי+ ההסתברות לקבל תוצאות קיצוניות יותר לכיוון H_1.</u>

כדי לחשב את ה- $\boldsymbol{P-Value}$ של המבחן מחלצים את אלפא מהמשוואה הבאה:

$\overline{X_n}=\mu_0\pm Z_{1-\alpha}\sqrt{\frac{\sigma^2}{n}}$ (מבחן חד צדדי), <u>**או:**</u>

$\overline{X_n}=\mu_0\pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}$ (מבחן דו צדדי)

← ערך ה- P של מבחן דו צדדי הוא פי 2 משל מבחן חד צדדי. מסתכלים על הצד המתאים לפי ממוצע המדגם שקיבלנו.



<u>הערות:</u>

1. $P-Value>\alpha$ - לא נדחה את H_0.
2. $P-Value<\alpha$ - נדחה את H_0.
3. תחת H_0, $P-Value$ (טעות מסוג ראשון) לא תלויה בגודל המדגם: $P_{H_0}(P-Value<\alpha)=\alpha$
4. תחת H_1, $P-Value$ הולך וקטן ככל שהמדגם גדל.

$Z_{0.9}$	$=1.285$
$Z_{0.95}$	$=1.645$
$Z_{0.975}$	$=1.96$
$Z_{0.99}$	$=2.325$

<u>הערות</u>

- אם דחיתי את H_0 בר"מ $\alpha = 0.05$ ← אדחה גם בר"מ $\alpha = 0.08$.
- אם דחיתי את H_0 בר"מ $\alpha = 0.08$ ← לא בהכרח אדחה גם בר"מ $\alpha = 0.05$:
  אם $C^* < \bar{x} < C$ - נדחה
  אם $\bar{x} < C^*$ - לא נדחה
- אם לא דחיתי את H_0 בר"מ $\alpha = 0.08$ ← לא אדחה גם בר"מ $\alpha = 0.05$.
- אם לא דחיתי את H_0 בר"מ $\alpha = 0.05$ ← לא בהכרח שלא אדחה גם בר"מ $\alpha = 0.08$:
  אם $C^* < \bar{x} < C$ - נדחה
  אם $\bar{x} < C^*$ - לא נדחה

<u>ככל ש- α גדלה, הערך הקריטי קטן, אבל אזור הדחייה גדל- אנו מוכנים לעשות יותר טעויות מסוג 1.</u>

<u>התפלגות χ^2 היא סימטרית וחיובית תמיד</u>

<u>הקשר בין רווח סמך לבדיקת השערות במבחן דו צדדי:</u>

אם μ_0 שייכת לרווח הסמך אז מתקיים :

$$\bar{x}_n - \mu_0 < Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}$$

ולכן לא נדחה את H_0 באותה רמת מובהקות.

<u>העוצמה היא פונקציה של μ האמיתי:</u>

<u>משפט הגבול המרכזי</u>		
<u>הסתכלות על ממוצעים</u>	**<u>הסתכלות על סכומים</u>**	
$\dfrac{\overline{x_n} - \mu}{\sqrt{\dfrac{\sigma^2}{n}}} \sim N(0,1)$	$\dfrac{\sum_{i=1}^{n} x_i - n\mu}{\sqrt{n\sigma^2}} \sim N(0,1)$	כל מ"מ פואסוני הוא סכום של מ"מ פואסוניים ולכן ניתן באמצעות משפט הגבול המרכזי של סכומים לחשב הסתברויות למשתנה זה

מבחנים א- פרמטריים

מבחן (χ^2) chi טיב התאמה

מבנה המבחן

H_0: המודל שלי

H_1: לא המודל שלי

סימונים

P_i	ההסתברויות לפי המודל של השערת האפס
O_i	תוצאות המדגם שלקחנו
$P_i \times N = E_i$	הציפייה לתוצאות המדגם על סמך המודל
N	גודל המדגם
n	מספר הקטגוריות
$n-1$	מספר דרגות החופש

איך זה נראה?

	קטגוריה 1	קטגוריה 2	קטגוריה 3	....	קטגוריה n
P_i					
O_i					
E_i					

מסקנה

דחה את H_0 אם:

$$\sum_{i=1}^{n}\frac{(O_i-E_i)^2}{E_i} > \chi^2_{n-1,\alpha}$$

הערה: אם P אינו ידוע- אי אפשר לבנות את E_i. על כן, נאמוד את P מהמדגם, כתוצאה מכך נאבד דרגת חופש, ומספר דרגות החופש יהיה $n-2$. למעשה, כל פרמטר שנאמד במדגם והוא לא בתוך הסטטיסטי גורר איבוד דרגת חופש.

מבחן הסימן

מבנה המבחן

$H_0: p=\frac{1}{2}$ ההקצאה מקרית

$$H_1:\begin{cases} p\neq\frac{1}{2} & \text{דו צדדי} \\ p<\frac{1}{2} & \text{חד צדדי שמאלי} \\ p>\frac{1}{2} & \text{חד צדדי ימני}\end{cases}$$

סימונים

$x_i - y_i$	תצפיות מזווגות בלי הנחת נורמאליות
$D_i = x_i - y_i$	סימן ההפרש (+/-)
n	מס' התצפיות המזווגות
p	פרופורציית ההפרשים החיוביים
Z	מס' ההפרשים החיוביים
k	מס' ההפרשים החיוביים שקיבלנו במדגם

איך זה נראה?

	תצפית מזווגת 1	תצפית מזווגת 2	תצפית מזווגת 3	....	תצפית מזווגת n
x_i					
y_i					
D_i					

$$p-value=\begin{cases} p(Z\le k)+p(Z\ge n-k) & \text{דו צדדי} \\ p(Z\le k) & \text{חד צדדי שמאלי} \\ p(Z\ge k) & \text{חד צדדי ימני}\end{cases}$$

מסקנה

כדי להגיע למסקנות עלינו ראשית לחשב $p-value$: ההסתברות לקבל תוצאה כפי שראינו במדגם או תוצאות קיצוניות יותר לכיוון H_1. לאחר החישוב הכלל הוא כזה:

דחה את H_0 אם: $\alpha < p-value$

הערה: כאשר מחשבים $p(Z=k)$ יש להתחשב בכל הסידורים האפשריים:

$$p(Z=k)=\binom{n}{k}\left(\frac{1}{2}\right)^{k}\left(\frac{1}{2}\right)^{n-k}$$

המבחן המדויק של פישר (טבלה 2X2)

מבנה המבחן

אנו מעוניינים להשוות בין הפרופורציות שתי אוכלוסיות על משתנה בינארי, למשל חולים/ בריאים ביחס למעשנים/ לא מעשנים.

$H_0: p_x = p_y$ הפרופורציות של שתי האוכלוסיות זהות.

$$H_1:\begin{cases} p_x\neq p_y \\ p_x>p_y \\ p_x<p_y\end{cases}$$

הפרופורציות אינן זהות או נוטות לאוכלוסיה ספציפית.

סימונים

a,b,c,d	תוצאות המדגם בכמויות ולא פרופורציות

$a+b+c+d=n$

סך התצפיות

איך זה נראה?

	תצפית מזווגת 1	תצפית מזווגת 2	תצפית מזווגת 3
אוכלוסייה א'	a	b	$a+b$
אוכלוסייה ב'	c	d	$c+d$
	$a+c$	$b+d$	n

כל הטבלה נקבעת בצורה חד חד ערכית ע"י תא אחד.

$$p(\text{לוח})=\frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}}$$

עבור תא a או c

מסקנה

ההחלטה במבחן תהיה על סמך חישוב $p-value$: **סכום כל הלוחות שיותר קיצוניים או שווים ללוח שראיתי.**

כדי לחשב את ה- $p-value$ אני אצור שורה של לוחות קיצוניים יותר לכיוון H_1 מנקודת המבט של תא a ביחס לתא השולי שלו $a+c$, וסכימה של כל ההסתברויות הללו הוא ה- $p-value$.

לאחר החישוב הכלל הוא כזה:

דחה את H_0 אם: $\alpha < p-value$

דגש: במבחן דו צדדי עלינו לקחת לוחות יותר קיצוניים משני הצדדים- את הערכה לקיצוניות נקבל על סמך ההסתברות של אותו לוח- האם היא קטנה או שווה ללוח שראיתי במדגם.