עידו שמיר אביב 2010, מרצה: פרופ' מלכה גורפיין

דף נוסחאות לקורס מבוא לסטטיסטיקה – 094423 – מבחן

שגיאת דגימה שגיאה כתוצאה מכך שנאספו נתונים על מדגם שהנו חלק מכלל האוכלוסייה. גודל השגיאה תלוי בשיטת הדגימה ובגודל המדגם.
שגיאה שאינה שגיאת דגימה נובעת משימוש בשיטת דגימה לא נכונה.
דגימה הסתברותית דגימה שבה ידועה ההסתברות של כל פרט האוכלוסייה להיבחר.
דגימה מקרית פשוטה דגימה הסתברותית כך שכל שתי קבוצות שוות גודל הן בעלות הסתברות שווה להיכלל במדגם. **הנחה** דגימה עם החזרה (לשם נוחות).

שונות $S^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2=\frac{1}{n-1}\sum_{i=1}^n(X_i^2-2X_i\bar{X}+\bar{X}^2)$ - מדד לפיזור סביב לממוצע.
חציון הערך שמחצית גדולים או שווים לו ומחצית קטנים או שווים לו.
אינו רגיש לערכים חריגים (בניגוד לממוצע). $S^2=\frac{1}{n-1}\left[\sum X_i^2-\frac{2}{n}(\sum X_i)^2+n(\bar{X})^2\right], E(S^2)=\sigma^2$
טווח $Range=X_{(n)}-X_{(1)}$
רבעון עליון הערך $\frac{1}{4}$ גדולים או שווים לו ו $\frac{3}{4}$ קטנים או שווים לו Q_3
רבעון תחתון הערך $\frac{3}{4}$ גדולים או שווים לו ו $\frac{1}{4}$ קטנים או שווים לו Q_1
תחום בין רבעוני התחום בו מצויות 50% מהתצפיות המרכזיות:
$IR=[Q_1,Q_3]$
מדד פיזור שאינו רגיש לערכים חריגים.
שיברון p באוכלוסיה ערך ε_p כך שמתקיים $P(X\le\varepsilon_p)=p$
שיברון p במדגם ערך x שפרופורציה p מתוך n ערכי המדגם קטנים-שווים לו.
- ככל שהמדגם גדל, יתקרב השיברון המדגמי לתיאורטי.
$step=\frac{3}{2}(Q_3-Q_1), LF=Q_1-step, UF=Q_3+step$
ה $whisker$ הוא התצפית הראשונה מכל כיוון ללא חריגה מ LF/UF

לכל מספר ממשי **פונקציית ההתפלגות האמפירית** מוגדרת:
$$F_n(x)=\frac{1}{n}\sum_{i=1}^n I(X_i\le x)$$
סטטיסטי הוא פונקציה של התצפיות במדגם – של המ"מ, ואינו תלוי בפרמטר לא ידוע. סטטיסטי המשמש לאמידת פרמטר נקרא **אומד** עבור הפרמטר. הערך הספציפי שהתקבל מהמדגם נקרא **אומדן**.

יהי X מ"מ. **המומנט ה- r** של X מוגדר $\mu_r=E(X^r)$ אם התוחלת מוגדרת.
יהי $X_1,\dots,X_n$ מדגם מקרי מהתפלגות $F_X(\cdot)$. **המומנט המדגמי ה-r** מוגדר:
$$\mu_1=E(X)$$
$$\mu_2=Var(X)+\mu_1^2$$
$$M_r=\frac{1}{n}\sum_{i=1}^n X_i^r$$
- $E(M_r)=\frac{1}{n}\sum_{i=1}^n E(X_i^r)=\mu_r$
- לפי חוק המספרים הגדולים $M_r\to\mu_r$
חיסרון: לא לוקחים בחשבון אילוצים על תחומים. **יתרון:** פשטות
- אומד בשיטת המומנטים עבור $\theta, Uni[0,\theta]$ הוא $\hat{\theta}=2\bar{X}$ (חסר הטיה, עקיב).
$$\sum_{i=1}^n(X_i-\bar{X})^2=\sum_{i=1}^n(X_i-\mu)^2-\sum_{i=1}^n(\bar{X}-\mu)^2$$

פונקצית הנראות של n מ"מ $X_1,\dots,X_n$ מוגדרת להיות הצפיפות (או ההסתברות) המשותפת: $L(\theta)=L(\theta;x_1,\dots,x_n)=f_{X_1,\dots,X_n}(x_1,\dots,x_n;\theta)$
אומד נראות מרבית (אנ"מ) הוא הערך θ המביא למקסימום את $L(\theta)$.
- אינו בהכרח יחיד
- קל לעבוד עם $\ln(\theta)$
- אם $\hat{\theta}$ אנ"מ ל θ אזי לכל פונקציה $\tau(\theta)$, אנ"מ לה הוא $\tau(\hat{\theta})$.
- אנ"מ ל p כאשר $\theta=p, Ber(p)$ הוא $\bar{X}$.
- אנ"מ עבור $\theta=(\mu,\sigma^2), N(\mu,\sigma^2)$ נקבל: $\hat{\mu}=\bar{X},\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2$.
- אנ"מ עבור $\theta, Uni[0,\theta]$ הוא $\hat{\theta}=X_{(n)}$ (מוטה, עקיב).
משופר: $\theta^*=\frac{n+1}{n}X_{(n)}$ (אח"ה, עקיב).

יהי $T_n=T(X_1,\dots,X_n)$ סטטיסטי המבוסס על מדגם בגודל n והנו אומד ל $\tau(\theta)$. **ההטיה** ($bias$) של T_n ביחס ל $\tau(\theta)$ מוגדרת על-ידי $b_{T_n}(\theta)=E(T_n)-\tau(\theta)$
נאמר כי אומד T_n הנו **אומד חסר הטיה (אח"ה)** אם $b_{T_n}(\theta)=0$ לכל $\theta\in\Theta$
- אם נוציא הרבה מאוד מדגמים, ועל-סמך כל מדגם נחשב את הערך של האומד, אזי הממוצע של כל האומדנים יהיה שווה בקירוב לערך הפרמטר אותו רוצים לאמוד.

השגיאה הריבועית הממוצעת (MSE) של אומד T_n ביחס ל $\tau(\theta)$ מוגדרת כ
$MSE_{T_n}(\theta)=E\left[\left(T_n-\tau(\theta)\right)^2\right]$
- אם $MSE=0$ אז בהכרח $T_n=\tau(\theta)$.
- אם $MSE_{T_n}(\theta)\le MSE_{S_n}(\theta)$ לכל $\theta\in\Theta$ אז נעדיף את T_n משיקולי MSE.
$$MSE(T)=E\left[\left(T_n-\tau(\theta)\right)^2\right]=Var(T)+\underbrace{\left(E(T)-\tau(\theta)\right)^2}_{\text{ריבוע ההטיה}}$$

אם $T_1,\dots,T_n$ סדרת אומדים ל $\tau(\theta)$ כך ש $T_n=T_n(X_1,\dots,X_n)$ נאמר כי סדרת האומדים **עקיבה** ל $\tau(\theta)$ אם לכל $\varepsilon>0$ מתקיים:
$$\lim_{n\to\infty}P(|T_n-\tau(\theta)|<\varepsilon)=1$$
משפט אם $MSE(T)\to 0$ אזי $\lim_{n\to\infty}MSE(T)\to 0$ הסדרה עקיבה ביחס ל $\tau(\theta)$.
- אם $\{T_n\}$ סדרת אומדים עקיבה ל θ ו $\tau(\theta)$ רציפה אזי $\{\tau(\theta)\}$ סדרת אומדים עקיבה ל $\tau(\theta)$.
- אם T_n אח"ה ל θ ו $\tau(\theta)$ ליניארית אז $\tau(T_n)$ אח"ה ל $\tau(\theta)$.
- מומנטים הם עקיבים וח"ה

סטטיסטי מספיק היא פונקציה של המדגם הנוצרת בתוכה את כל האינפורמציה שיש במדגם ביחס לפרמטר.
סטטיסטי $S=S(X_1,\dots,X_n)$ יקרא **סטטיסטי מספיק** אם"ם ההתפלגות המותנה של $X_1,\dots,X_n$ בהינתן $S=s$ ב"ת ב θ לכל ערך s.
- ס"מ אינו יחיד וכל פונקציה חח"ע של ס"מ היא גם ס"מ.
$T^*(X_1,\dots,X_n)=r\left(T(X_1,\dots,x_n)\right)$
עבור $X_1,\dots,X_n\sim Ber(p)$ ב"ת $\sum_{i=1}^n X_i$ ס"מ ל p וגם $\bar{X}$.
עבור $X_1,\dots,X_n\sim N(\mu,\sigma^2)$ ב"ת $\left(\sum_{i=1}^n X_i^2,\sum_{i=1}^n X_i\right)$ ס"מ ל $\theta=(\mu,\sigma^2)$
וגם $\left(\bar{X},\sum_{i=1}^n X_i^2\right)$ ס"מ
אם השונות אינה ידועה והתוחלת ידועה אזי ס"מ לשונות הוא $\sum_{i=1}^n(X_i-\mu)^2$
אם התוחלת אינה ידועה והשונות ידועה, אזי ס"מ לתוחלת הוא $\bar{X}$
משפט אנ"מ תלוי במדגם מקרי דרך ס"מ

משפט ראו-בלקוול אומד שהנו פונקציה של ס"מ הוא עם MSE קטן יותר בהשוואה ל MSE של אומד שאינו פונקציה של ס"מ.
יהי $X_1,\dots,X_n$ מדגם מקרי ויהי S ס"מ. T אומד ל $\tau(\theta)$ נגדיר $T^*=E(T|S)$:
1. T^* הינו סטטיסטי ופונקציה של S
2. $E(T^*)=E(T)$
3. לכל $\theta\in\Theta$: $MSE_{T^*}(\theta)\le MSE_T(\theta)$
*** לא ניתן לשפר שוב**

רווח סמך: **רווח סמך קטן הוא יותר מדויק (יותר נתונים מקטינים את ר"ס)**
משמעות: לר"ס ברמת סמך α – 1 עבור פרמטר כלשהו, יש הסתברות של α – 1 שהרווח שיתקבל ע"ס מדגם מקרי יכלול בתוכו את הערך האמיתי של הפרמטר.
התפלגות נורמאלית $N(\mu,\sigma^2)$
ר"ס לתוחלת, שונות ידועה: $\bar{X}\pm Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$
- סימטרי סביב ממוצע המדגם, מיקומו אקראי
- רווח-סמך הקצר ביותר
- אורכו קבוע מראש ושווה ל $2Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$
אם נחפש **שאורך רווח הסמך** לא יעלה על $2d$: $2Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\le 2d\to n\ge\frac{Z_{1-\frac{\alpha}{2}}^2\sigma^2}{d^2}$
בניית רווח-סמך כללי:
1. אנ"מ לפרמטר θ
2. בחירת פונקציה פיבוטלית $Q(\theta,\hat{\theta})$ שהתפלגותה ידועה $(f,\chi^2,t,N(0,1))$ אך אינה תלויה בפרמטר θ
3. מציאת השיברונים X_{p_1},X_{p_2} כך ש $p_2-p_1=1-\alpha$
4. $P(X_{p_1}\le Q\le X_{p_2})=1-\alpha$
- את $\hat{\theta}$ נבחר כך שיהיה פונקציה של ס"מ.
- אם נרצה רווח-סמך חד-צדדי נבחר $p_1=0$ או $p_2=1$
ר"ס לתוחלת, שונות איננה ידועה: $\bar{X}\pm\frac{S}{\sqrt{n}}t_{(n-1),1-\frac{\alpha}{2}}$
- רווח-סמך הקצר ביותר
- מרכז רווח הסמך הוא ממוצע המדגם
- אורך רווח הסמך הוא משתנה מקרי: $2t_{(n-1),1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$
לכן אין אפשרות לחשב מראש את גודל המדגם שיבטיח רווח באורך רצוי מאחר ואורך הרווח תלוי באומד לסטיית התקן שיתקבל במדגם.
רווח סמך לפרופורציה $X\sim Bin(n,p)$ רווח סמך עבור p ברמת סמך α – 1 בקירוב: $\hat{p}\pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- מבוסס על קירובים: משפט הגבול המרכזי ואמידת השונות.
אורך רווח הסמך: $2Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ ולכן מספר התצפיות הדרוש כדי שהסטייה לא תעלה על גודל d מסוים ברמת הביטחון הנתונה היא: $n\ge\frac{Z_{1-\frac{\alpha}{2}}^2\hat{p}(1-\hat{p})}{d^2}$
גודל המדגם הגרוע ביותר: $\hat{p}(1-\hat{p})\le 0.25$
ר"ס לשונות, תוחלת ידועה: $\left[\frac{\sum_{i=1}^n(X_i-\mu)^2}{\chi^2_{(n),1-\frac{\alpha}{2}}},\frac{\sum_{i=1}^n(X_i-\mu)^2}{\chi^2_{(n),\frac{\alpha}{2}}}\right]$
ר"ס לשונות, תוחלת איננה ידועה: $\left[\frac{(n-1)S^2}{\chi^2_{(n-1),1-\frac{\alpha}{2}}},\frac{(n-1)S^2}{\chi^2_{(n-1),\frac{\alpha}{2}}}\right]$
- איננו האופטימאלי מבחינת אורכו
- עבור סטיית התקן – פשוט שורש
רווח סמך להפרש תוחלות $\mu_x-\mu_y$: $(\bar{X}-\bar{Y})\pm Z_{1-\frac{\alpha}{2}}\sqrt{\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)}$
אם השונות אינה ידועה (שוניות שוות): $S_P^2=\frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}$ וכן
$\frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sqrt{S_P^2\left(\frac{1}{n}+\frac{1}{m}\right)}}\sim t_{(n+m-2)}$ ולכן $(\bar{X}-\bar{Y})\pm t_{(n+m-2),1-\frac{\alpha}{2}}\sqrt{S_P^2\left(\frac{1}{n}+\frac{1}{m}\right)}$
שוניות לא שוות: $Q=\frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}}\sim t_{(df)}$ כאשר $df=\frac{\left(\frac{S_X^2}{n}+\frac{S_Y^2}{m}\right)^2}{\frac{S_X^4}{n^2(n-1)}+\frac{S_Y^4}{m^2(m-1)}}$ ולכן ר"ס
$(\bar{X}-\bar{Y})\pm t_{(df),1-\frac{\alpha}{2}}\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}$
ר"ס ליחס שוניות $\lambda=\frac{\sigma_X^2}{\sigma_Y^2}$: היות ש $\frac{S_X^2}{S_Y^2}\frac{1}{\lambda}\sim F_{(n-1,m-1)}$ אז ר"ס הוא
$\left[\frac{S_X^2}{S_Y^2}F_{(m-1,n-1)}^{\frac{\alpha}{2}},\frac{S_X^2}{S_Y^2}F_{(m-1,n-1)}^{1-\frac{\alpha}{2}}\right]$
התפלגות ברנולי $Ber(p_x),Ber(p_y)$
ר"ס ל p_x-p_y: $(\hat{p}_x-\hat{p}_y)\pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n}+\frac{\hat{p}_y(1-\hat{p}_y)}{m}}$
התפלגויות מזווגות $(X_1,Y_1),(X_2,Y_2),\dots,(X_n,Y_n)$
ר"ס להפרש התוחלת $\mu_x-\mu_y$:
נגדיר $D_1,\dots,D_n$ אזי $D_i=X_i-Y_i$ ב"ת ושווי התפלגות וכן:
$E(D_i)=E(X_i)-E(Y_i)=\mu_x-\mu_y\equiv\mu_d$
נניח $D_i\sim N(\mu_D,\sigma_D^2)$ ולכן: $\bar{D}\pm t_{(n-1),1-\frac{\alpha}{2}}\frac{S_D}{\sqrt{n}}$
- לא ניתן להתעלם מהתלות
- רווח הסמך השגוי (אם נתעלם מהתלות) יהיה רחב יותר מרווח הסמך המזווג כי $Var(\bar{X}-\bar{Y})=Var(\bar{X})+Var(\bar{Y})-2Cov(\bar{X},\bar{Y})$ ו $2Cov(\bar{X},\bar{Y})>0$

סימולציה ליצירת $X\sim\exp(\lambda)$
$$U=F_X(x)=1-e^{-\lambda x}\to F^{-1}(u)=\frac{\ln(1-u)}{-\lambda}\to X=-\frac{\ln(1-U)}{\lambda}$$
התפלגות אחידה $X\sim Uni(a,b)$
$$P(c\le x\le d)=\frac{d-c}{b-a},\qquad f_X(x)=\begin{cases}\frac{1}{b-a} & a\le x\le b\\ 0 & ,else\end{cases}$$
$$E(X)=\frac{a+b}{2},Var(X)=\frac{(b-a)^2}{12}$$
$$\frac{X-a}{b-a}\sim Uni(0,1)$$

פונקציה יוצרת מומנטים:
$$M_X(t)=E(^{tX})$$
$$iid: M_{X+Y}=M_X\cdot M_Y$$
$$M_X(0)=1$$
$$M_{aX+b}(t)=e^{bt}M_X(at)$$

משפט הפירוק סטטיסטי $S(x_1,\dots,x_n)$ הינו מספיק ביחס לפרמטר θ אם"ם לכל $x_1,\dots,x_n$ מתקיים (צפיפות או הסתברות):

$$f_{x_1,\dots,x_n}(x_1,\dots,x_n;\theta)=h(x_1,\dots,x_n)\cdot g(S(x_1,\dots,x_n);\theta)$$

כאשר h פונקציה אי-שלילית ואינה פונקציה של θ. ו g אי-שלילית ותלויה במשתנים שלה רק דרך S.

<u>טרנספורמציה חד-חד-ערכית של מ"מ רציף:</u>

X מ"מ רציף (דיפרנציאבילי) המקבל ערכים בקטע $[a,b]$ ו $g:\mathbb{R}\to\mathbb{R}$ פונקציה גזירה ועולה (יורדת) ממש, $Y=g(x)$:

$$F_y(y)=\begin{cases}f_x(g^{-1}(y))\cdot\left|\frac{\partial}{\partial y}g^{-1}(y)\right| & ,g(a)\le y\le g(b)\\ 0 & ,else\end{cases}$$

$$Y=a+bX\to f_Y(y)=\frac{1}{|b|}f_x\left(\frac{y-a}{b}\right)$$

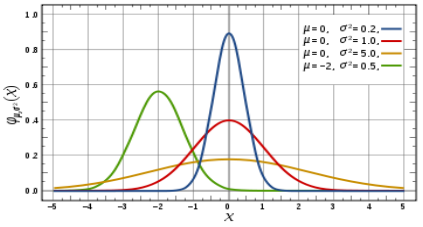
התפלגות נורמלית $X\sim N(\mu,\sigma^2)$

$$E(X)=\mu\,,Var(X)=\sigma^2$$

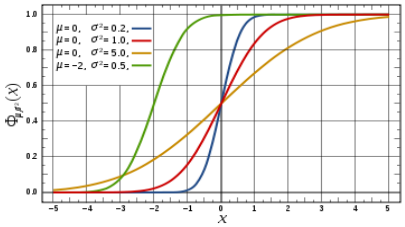
$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\quad,\quad \bar X\sim N\left(\mu,\frac{\sigma^2}{n}\right)$$

כל קומבינציה ליניארית של משתנים נורמליים הוא נורמלי.

פונקצית הצפיפות:



פונקציית הצטברות:



אקספוננציאלי מוזז

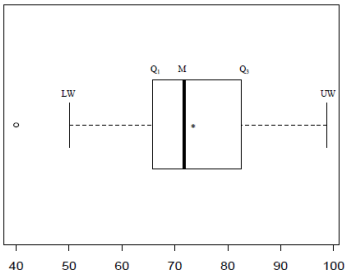
$$X=Y+\theta\sim\exp(\lambda)$$

$$f(x)=\lambda e^{-(x-\theta)\lambda}$$

סיכום קטן לאמדים

	מומנטים□אמת	אנ"מ	ס"מ
N	$\hat\mu=\bar X$ $\hat\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar X)^2$	$\hat\mu=\bar X$ $\hat\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar X)^2$	$(\bar X,\sum X_i^2)$ ס"מ ל $\theta=(\mu,\sigma^2)$
$Pois$	$\hat\lambda=\bar X$	$\lambda=\bar X$	$\sum X_i$
exp	$\hat\theta=\frac{1}{\bar X}$	$\hat\theta=\frac{1}{\bar X}$	$\sum X_i$
Bin	$\hat p=\bar X$		
Ber	$\hat p=\bar X$	$\hat p=\bar X$	$\sum X_i$

דיאגרמת קופסא



- בדיאגרמת קופסא ניתן לראות את מרבית הסטטיסטים: חציון, רבעון עליון ותחתון, *Whiskers*, טווח הנתונים וטווח בין-רבעוני.
- יש תצפית אחת הקטנה מהגדר התחתונה ולכן זאת תצפית חרגיה.
- ההתפלגות אינה סימטרית. חוסר הסימטריה מתבטא בעיקר במחצית המרכזית של הנתונים, כלומר בקופסא. הממוצע גבוה מהחציון, ומרחקו של הרבעון העליון מהחציון גדול יותר מאשר מרחקו של הרבעון התחתון מהחציון. להתפלגות נטייה ימינה.
- ניתן ללמוד על פיזור הנתונים סביב הממוצע והחציון.

Pvalue הנה ההסתברות לקבל תוצאה קיצונית לפחות כמו התוצאה שקיבלנו בניסוי, בהנחה שהשערת האפס נכונה.

אם $Pvalue\ge\alpha$ נדחה את השערת האפס בר"מ α

אחרת לא נדחה את השערת האפס בר"מ α

הגדרה 1 ההסתברות לקבל תוצאה מדגמית כפי שקיבלנו או קיצונית ממנה בכיוון H_1.

הגדרה 2 α <u>מקסימאלית</u> עבורה עדיין מקבלים את השערת H_0.

או α <u>מינימאלית</u> שממנה מתחילים לדחות את H_0.

מקדם המתאם הלינארי בין שני משתנים מקריים מוגדר $\rho=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$ $(-1\le\rho\le1)$

$$Cov(X,Y)=E[\{X-E(X)\}\{Y-E(Y)\}]=E(XY)-E(X)E(Y)$$

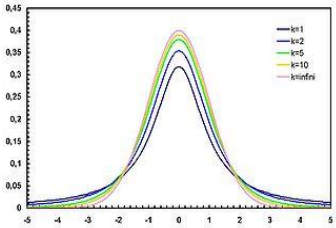
האומד למקדם המתאם הלינארי (מדגמי) מוגדר להיות $R=\frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}=\hat\beta_1\sqrt{\frac{S_{XX}}{S_{YY}}}$.

- הערך של R אינו תלוי באיזה משתנה נקרא Y ואיזה X.
- הערך של R אינו תלוי ביחידות המדידה של המשתנים.
- תמיד בין מינוס 1 ל1. אם הוא קרוב ל0 אז הוא מעיד על חוסר קשר **ליניארי** בין המשתנים.

התפלגות t נראה מאוד דומה לנורמלי. יהי $Z\sim N(0,1)$ ו $W_k\sim\chi^2_{(k)}$ ב"ת אז

$$T=\frac{Z}{\sqrt{\frac{W_k}{k}}}\sim t_{(k)}\ ,\qquad f_T(t)=\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}}\left(1+\frac{t^2}{k}\right)^{\frac{k+1}{2}}\ ,k>0,-\infty<t<\infty$$

$$E(T)=0\ ,k>1\ \ ,Var(T)=\frac{k}{(k-2)}\ ,k>2$$



- סימטרי סביב 0
- פי"מ לא קיימת
- זנבות עבים יותר משל נורמלי

$\Gamma(n)=(n-1)!$

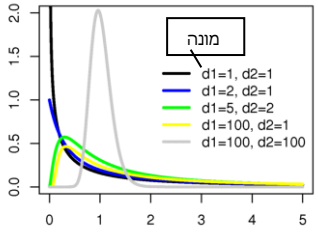
$\Gamma(0.5)=\sqrt{\pi}$

יהיו $X_1,\dots,X_n\sim N(\mu,\sigma^2)$ אז מתקיים $T=\frac{\bar X-\mu}{\frac{S}{\sqrt{n}}}\sim t_{(n-1)}$

התפלגות F יהיו $U_1\sim\chi^2_{(k)}$ ו $U_2\sim\chi^2_{(m)}$ ב"ת

$$F=\frac{\frac{U_1}{k}}{\frac{U_2}{m}}\sim F_{(k,m)}\ ,\qquad f_F(y)=\frac{\Gamma\left(\frac{k+m}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}\left(\frac{k}{m}\right)^{\frac{k}{2}}\frac{y^{\frac{k}{2}-1}}{\left(1+\frac{ky}{m}\right)^{\frac{k+m}{2}}}\ ,y>0$$

$$E(F)=\frac{m}{m-2}\ ,m>2\ ,Var(F)=\frac{2m^2(k+m-2)}{k(m-2)^2(m-4)}\ ,m>4$$



- $\lim_{m\to\infty}kF\sim\chi^2_{(k)}$
- $\frac{1}{F}\sim F_{(m,k)}$
- אם $X\sim t_{(d)}$ אז $X^2\sim F_{(1,d)}$
- אם $P(Y\le q_Y(p))=p$ ו $Y=\frac{1}{F}$ אז $P(Y\le q_Y(1-p))=1-p$ ו $q_X(p)=\frac{1}{q_Y(1-p)}$

הסתברות חשובות

$$\mathrm{Var(X)=E(X^2)-\left(E(X)\right)^2}$$

$$Cov(X,Y+Z)=Cov(X,Y)+Cov(X,Z)$$

$$Var(Y)=E(Var(Y|X))+Var(E(Y|X))$$

אם $X_i\sim Pois(\lambda)$ אז $P(\sum X_i=s)=\frac{(n\lambda)^s e^{-n\lambda}}{s!}$ וגם $(X\_k|\sum X_i=s)\sim Bin\left(s,\frac{1}{n}\right)$

חיבור פואסון $X+Y\sim Pois(\mu+\lambda)$.

מג"מ $X_1,\dots,X_n$ מ"מ ב"ת ש"ה $E(X_i)=\mu$,$Var(X_i)=\sigma^2$ אז $\frac{\bar X_n-\mu}{\frac{\sigma}{\sqrt{n}}}\to N(0,1)$

WALD אם X_i ב"ת ש"ה ו N בדיד אז $E(X_1+\dots+X_N)=E(X_i)\cdot E(N)$.

ב"ת גורר $Cov(X,Y)=0$ וגם $Cov(X,X)=Var(X)$

אינדיקטורים: $E(I)=E(I^2)=p$ וגם $Var(I)=pq$

מרקוב: $P(X\ge a)\le\frac{E(X)}{a}$

צ'בשב: X מ"מ כך ש $E(X)=\mu$ ו $Var(X)=\sigma^2$ אז $P(|X-\mu|\ge a)<\frac{Var(X)}{a^2}$

מינימום של אקספוננט הוא אקספוננט $Y\sim\exp(\mu)$, $X\sim\exp(\lambda)$ אז:

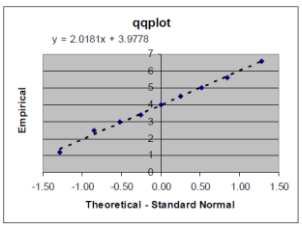
$$\min(X,Y)\sim\exp(\lambda+\mu)$$

וגם **תחרות בין אקספוננטים** $P(X<Y)=\frac{\lambda}{\lambda+\mu}$

סכום גאמות אם $X\sim Gamma(r,\lambda)$ ו $Y\sim Gamma(s,\lambda)$ אז $X+Y\sim Gamma(r+s,\lambda)$.

אקספוננציאלי $F_X(a)=P(X\le a)=1-e^{-\lambda a}$

QQPLOT



i	1	2	3	4
$\frac{i}{n+1}$	0.1	0.2	0.3	0.4
שבחונים תיאורטיים עפ"י התפלגות נורמלית סטנדרטית	-1.282	-0.842	-0.524	-0.253
שבחונים אימפירים (התצפיות מסודרות מהקטנה לגדולה)	1.2	2.5	3	3.4

<u>דוגמא</u>

מדגם מקרי $X_1,\dots,X_n$ מהתפלגות $Ber(p)$

נתעניין בבדיקת השערות

$H_0:p=p_0\qquad H_1:p=p_1$

מבחן "נ": נדחה את השערת האפס אם

$$\lambda(x)=\frac{\binom{n}{x}p_1^x(1-p_1)^{n-x}}{\binom{n}{x}p_0^x(1-p_0)^{n-x}}=\frac{p_1^x(1-p_1)^n(1-p_0)^x}{p_0^x(1-p_0)^n(1-p_1)^x}=\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)^x\left(\frac{1-p_1}{1-p_0}\right)^n>k_\alpha$$

$$\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)^x>k_\alpha\left(\frac{1-p_1}{1-p_0}\right)^{-n}=k_\alpha^*$$

אם $p_1>p_0$ אזי, $x\ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)>\ln k_\alpha^*=k_\alpha^{**}$

לכן, נדחה את השערת האפס אם $X=\sum_{i=1}^n X_i>C_\alpha$

אם $p_1<p_0$ אזי, $x\ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)<\ln k_\alpha^*=k_\alpha^{**}$

לכן, נדחה את השערת האפס אם $X=\sum_{i=1}^n X_i<C_\alpha$

נמצא את הערך של C_α:

נניח $p_1>p_0$ אזי המבחן הינו

דחה את השערת האפס אם $\sum_{i=1}^n X_i>C_\alpha$

ודחה את השערת האפס בהסתברות γ_α אם $\sum_{i=1}^n X_i=C_\alpha$

נניח $p_1=3/4\quad p_0=1/2\quad n=10\quad \alpha=0.05$

אזי

$$P_{H_0}(X=10)+P_{H_0}(X=9)=0.5^{10}+10\cdot0.5^{10}=0.01074<0.05$$

התפלגות חי-בריבוע

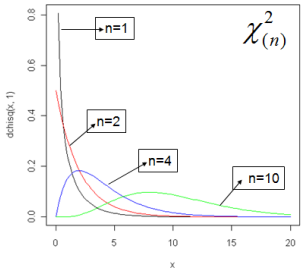
יהיו $Z_1,\dots,Z_n \sim N(0,1)$ מ"מ ב"ת ש"ה אזי: $W_n=\sum_{i=1}^n Z_i^2 \sim \chi^2_{(n)}$

- מקבל ערכים חיוביים בלבד
- צפיפות לא סימטרית עם זנב ימני
- עבור $n>50$ ההתפלגות קרובה מאוד לנורמלית
- סכום של מ"מ ב"ת המתפלגים חי-בריבוע, מתפלג חי-בריבוע

אם $X_1,\dots,X_n \sim N(\mu,\sigma^2)$ מ"מ ב"ת אז $W_n=\frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\mu)^2 \sim \chi^2_{(n)}$

אם התוחלת איננה ידועה אז $\frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\bar X_n)^2 \sim \chi^2_{(n-1)}$

היות ו $X_1-\bar X_n$ ו $\bar X_n$ **בלתי מתואמים** ומתפלגים נורמאלית, הם גם ב"ת ולפיכך $\bar X_n, \sum_{i=1}^n (X_i-\bar X_n)^2$ **בלתי תלויים**.



$$f_X(x)=\frac{1}{\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{2}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1}e^{-\frac{1}{2}x}I_{(0,\infty)}$$

$$E(X)=n\ ,\quad Var(X)=2n\ ,\quad E(e^{tX})=\left(\frac{1}{1-2t}\right)^{\frac{n}{2}},\ t<\frac{1}{2}$$

אם ידוע כי $Z_i \sim \exp(\lambda)$ מ"מ ב"ת אז $\sum_{i=1}^n Z_i \sim Gamma(n,\lambda)$ וגם:

$$2\lambda\sum_{i=1}^n Z_i \sim Gamma(n,\lambda)\sim Gamma\left(n,\frac{1}{2}\right)\sim\chi^2_{(2n)}\ ,\qquad 2\sum_{i=1}^n Z_i \sim Gamma\left(n,\frac{\lambda}{2}\right)$$

בדיקת השערות

ההחלטה

האמת		אי דחיית H_0	דחיית H_0
	H_0 נכונה	V	טעות מסוג I
	H_1 נכונה	טעות מסוג II	V

טעות מסוג I: דחיית השערת האפס כאשר השערת האפס נכונה.
טעות מסוג II: אי-דחיית השערת האפס כאשר השערת האפס אינה נכונה.

$$\alpha=P(type\ I\ error)=P_{H_0}(reject\ H_0)$$
$$\beta=P(type\ II\ error)=P_{H_1}(accept\ H_0)$$

בד"כ הקטנת ההסתברות לטעות מסוג ראשון גוררת הגדלת ההסתברות לטעות מסוג שני וההיפך.

השערה פשוטה היא השערה הקובעת באופן יחיד את ההתפלגות של האוכלוסייה שממנה נלקח המדגם. בכל מקרה אחר, ההשערה נקראת **השערה מורכבת** (מורכבת = לא שווה, גדול מ, קטן מ).

פונקציית המבחן לכל כלל הכרעה D מותאמת פונקציה $\phi_D(x)=\phi_D(x_1,\dots,X_n)$ (נקראת המבחן) $x\in X$ (אוסף התוצאות האפשריות) כך ש $0\le\phi_D(x)\le 1$ קובעת את ההסתברות בה נדחה את השערות האפס עבור התוצאה x.

$$\alpha=E_{H_0}(\phi_D(X)),\beta=E_{H_1}(1-\phi_D(X))$$

רמת המובהקות (ר"מ) של המבחן היא ההסתברות המירבית המותרת לטעות מסוג ראשון. נקבעת לפי מידת הנזק – נזק רב גורר ר"מ נמוכה.

העוצמה של המבחן מוגדרת להיות ההסתברות לדחיית השערת האפס כאשר האלטרנטיבה נכונה. $\pi=1-\beta$.

יחס נראות $\lambda(x)=\frac{f_1(x)}{f_0(x)}$ (צפיפות או הסתברות).

הלמה של $Neyman-Pearson$ **(מבחן יחס-נראות)** לבדיקת ההשערות הפשוטות

$H_0: X\sim f_0$
$H_1: X\sim f_1$

ברמת מובהקות הקטנה או שווה ל α, מבחן בעל עוצמה מירבית הינו

פונקציה יורדת של x
נותן מבחן $x<C$

עבור k_α אי-שלילי, $0\le\tau_\alpha\le 1$, וכן k_α ו τ_α נקבעים כך שיתקיים:

$$\alpha=P_{H_0}(reject\ H_0)=E_{H_0}(\phi(X))$$

- מבין כל המבחנים עם ר"מ α (או קטנים יותר), מבחן י"נ הוא מבחן עם עוצמה מירבית.

טענה אם $f_0(\cdot)=f(\cdot,\theta_0)$ ו $f_1(\cdot)=f(\cdot,\theta_1)$ אזי מבחן י"נ תלוי ב X רק דרך ס"מ.

תהי A משפחה של התפלגויות של X ועלינו לבדוק את ההשערות

$H_0: F=F_0$
$H_1: F\in A$

אם לכל $F_1\in A$ המבחן בעל עוצמה מירבית בר"מ α של F_0 מול F_1 הוא אותו מבחן, אזי נאמר כי מבחן זה הוא בעל **עוצמה מירבית במידה שווה** ביחס ל A, בר"מ α.

- $X_1,\dots,X_n\sim N(\mu,\sigma^2)$ מ"מ ב"ת, השונות ידועה. $H_0:\mu=\mu_0, H_1:\mu>\mu_0$ עבור $\mu_1>\mu_0$ מסוים, מבחן בר"מ α בעל עוצמה מירבית הינו:

דחה את השערת האפס אם $\bar X>C_\alpha=z_{1-\alpha}\frac{\sigma}{\sqrt n}+\mu_0$

(אינו תלוי בערך של $\mu_1>\mu_0$ ולכן זהו מבחן בעל עוצמה מירבית במידה שווה בר"מ α).

- אם נשתמש במ"מ המבחן לא יהיה בעל עוצמה מירבית.

הקשר בין שני הסטטיסטים. בעזרת $\log(1+\varepsilon)=\varepsilon-\frac{\varepsilon^2}{2!}+\frac{\varepsilon^3}{3!}-\cdots(-1)^{n+1}\frac{\varepsilon^n}{n!}$

$$\Lambda^*(X)=-2\sum_{j=1}^K X_j\log\left(\underbrace{\frac{np_{0j}}{X_j}-1}_{\varepsilon}+1\right)=-2\sum_{j=1}^K X_j\left\{\left(\frac{np_{0j}}{X_j}-1\right)-\frac{1}{2}\left(\frac{np_{0j}}{X_j}-1\right)^2+\cdots\right\}$$

$$\approx -2\sum_{j=1}^K\left(np_{0j}-X_j\right)+\sum_{j=1}^K\frac{1}{X_j}\left(np_{0j}-X_j\right)^2$$

$$\approx\sum_{j=1}^K\frac{1}{np_{0j}}\left(np_{0j}-X_j\right)^2=X_P^2$$

$=0$ $\quad E_{H_0}(X_j)=np_{0j}$

תוצאות בדיקת השערות לדוגמא

מדגם מקרי $X_1,\dots,X_n$ מהתפלגות $N(\mu,\sigma^2)$ והשונות ידועה.

$H_0:\mu=\mu_0$
$H_1:\mu=\mu_1$

אם $\mu_0<\mu_1$ דחה אם $\bar X>C_\alpha$:

$$C_\alpha=\frac{\sigma}{\sqrt n}z_{1-\alpha}+\mu_0$$

$$\pi=\Phi\left(\frac{\mu_1-\mu_0}{\frac{\sigma}{\sqrt n}}-z_{1-\alpha}\right)$$

- פונקציית העוצמה עולה ב $\mu_1-\mu_0$
- פונקציית העוצמה עולה ב n
- עבור $\mu_1=\mu_0$ נקבל $\pi=\alpha$
- גודל המדגם המינימאלי הדרוש לעוצמה ור"מ נתונות: $n=\frac{(z_\pi+z_{1-\alpha})^2\sigma^2}{(\mu_1-\mu_0)^2}$

לסיכום נדחה אם $\bar X<\mu_0-z_{1-\alpha}\frac{\sigma}{\sqrt n}$.

אם $\mu_1>\mu_0$ דחה אם $\bar X<C_\alpha$, זאת אומרת $\frac{\bar X-\mu_0}{\frac{\sigma}{\sqrt n}}<-z_{1-\alpha}=z_\alpha$

בדיקת ההשערות $H_0:\theta\in W, H_1\theta\in W^C$ כך ש $W\cup W^C=\Theta$

יחס נראות מוכלל מוגדר להיות:

$$\Lambda(x)=\frac{\overbrace{\sup_{\theta\in W} f_\theta(x)}^{H_0\text{ תחת}}}{\underbrace{\sup_{\theta\in\Theta} f_\theta(x)}_{H_0\cup H_1\text{ תחת}}}$$

אזי המבחן יהיה: דחה את השערת האפס אם $\Lambda<k_\alpha$ כאשר k_α נקבע כך שלמבחן תהיה ר"מ α.

- $\Lambda(X)$ נקרא סטטיסטי המבחן.
- $0\le\Lambda(X)\le 1$
- בד"כ מבחן טוב, אך יש מקרים בהם יש טובים ממנו.
- אם $X_1,\dots,X_n$ מדגם מקרי, אזי $\Lambda^*=-2\log\Lambda\sim_{n\to\infty}\chi^2_{(k-1)}$

מוכלל זהו מבחן יחס נראות כאשר יש צורך לאמוד פרמטרים.
במבחן יחס נראות מוכלל המונה והמכנה מתחלפים, וגם הסימן.

טיב התאמה

טיב התאמה למשתנה K קטגוריות (ערכים) $A_1,\dots,A_K$

בהסתברויות $p_1,\dots,p_K$ בהתאמה $(\sum_{i=1}^K p_i=1)$.

יש n תצפיות, כל תצפית מסווגות לאחת הקטגוריות.

X_j מספר התצפיות בקטגוריה j אזי $X_1,\dots X_K\sim Multi(n,p_1,\dots,p_K)$ מולטינומי.

$$\sum_{j=1}^n x_j=n$$

$$\hat p_j=\frac{X_j}{n}\ ,j=1,\dots,k$$

לבדיקת ההשערות:

$H_0:\vec p=\vec p_0$
$H_1:\vec p\ne\vec p_0$

כאשר $\vec p_0=(p_{10},\dots,p_{K0})$ ידועות

וכן $\sum_{j=1}^K p_{j0}=1$ תחת השערת האפס:

$$E(X_j)=np_{j0}$$

קטגוריה	A_1	A_2	$\cdots$	A_K	סה"כ
שכיחות במדגם	X_1	X_2	$\cdots$	X_K	n
שכיחות צפויה	np_{10}	np_{20}	$\cdots$	np_{K0}	n

$$\sum_{j=1}^K E(X_j)=\sum_{j=1}^K np_{j0}=n\sum_{j=1}^K p_{j0}=n$$

סטטיסטי המבחן **לבדיקת טיב ההתאמה** של המדגם למודל ההסתברותי מוגדר להיות:

$$X^2_{\vec p}=\sum_{j=1}^K\frac{\left(X_j-np_{j0}\right)^2}{np_{j0}}$$

טענה עבור מדגם "מספיק גדול" אם השערת האפס נכונה, אזי:

$$X^2_{\vec p}\sim\chi^2_{(K-1)}\ ,\qquad \left[X_j\sim_{H_0}Pois\left(np_{j0}\right)\right]$$

כאשר $K-1$ בגלל שהשכיחות בקטגוריה האחרונה נקבעת ע"ס כל הקודמות לה.
מקובל להניח שהקירוב לעיל מספיק טוב אם $p_{j0}\ge 5$ לכל j.
אם כלל האצבע לא מתקיים, נאחד קטגוריות סמוכות.
ניתן להשתמש ב $\Lambda^*=-2\log\Lambda\sim_{n\to\infty}\chi^2_{(k-1)}$.

הערה אם לצורך חישוב ההסתברויות תחת השערת האפס יש לאמוד r פרמטרים, אזי נאמד כל פרמטר על ידי אנ"מ וד"ח יהיו $K-1-r$.
כל שימוש באנ"מ מוריד דרגת חופש.

אמידה

$$E(Y|X=x_0)=\beta_0+\beta_1x_0\equiv\eta(x_0)$$

$$\hat\eta(x_0)=\hat\beta_0+\hat\beta_1x_0\sim N\left(\beta_0+\beta_1x_0,\sigma^2\left[\frac{1}{n}+\frac{(x_0-\bar X)^2}{S_{XX}}\right]\right)$$

מאחר ו $(\hat\beta_0,\hat\beta_1)$ ב"ת ב $\hat\sigma^2$ אזי

$$\frac{\hat\beta_0+\hat\beta_1x_0-(\beta_0+\beta_1x_0)}{\hat\sigma\sqrt{\frac{1}{n}+\frac{(x_0-\bar X)^2}{S_{XX}}}}\sim t_{(n-2)}$$

ולכן ר"ס עבור $\eta(x_0)$ (הקצר ביותר יתקבל בנקודה $\bar X$) הינו:

$$\hat\beta_0+\hat\beta_1x_0\pm\hat\sigma\sqrt{\frac{1}{n}+\frac{(x_0-\bar X)^2}{S_{XX}}}\cdot t_{\left(n-2,1-\frac{\alpha}{2}\right)}$$

ניבוי/תחזית נקבל תוספות בשונות עבור התצפית החדשה ולכן ר"ס:

$$\hat\beta_0+\hat\beta_1x_0\pm\hat\sigma\sqrt{\mathbf{1}+\frac{1}{n}+\frac{(x_0-\bar X)^2}{S_{XX}}}\cdot t_{\left(n-2,1-\frac{\alpha}{2}\right)}$$

רווח סמך יותר, והאורך **אינו** שואף לאפס כאשר גודל המדגם שואף לאינסוף.

רווח סמך ל σ^2: $\sigma^2\in\left[\frac{\hat\sigma^2(n-2)}{\chi^2_{\left(n-2,1-\frac{\alpha}{2}\right)}},\frac{\hat\sigma^2(n-2)}{\chi^2_{\left(n-2,\frac{\alpha}{2}\right)}}\right]$

	SS	DF	MS	F
Model	$SSR=\hat\beta_1^2S_{XX}=\hat\beta^T(X^TY)-n(\bar Y)^2$	r	$\frac{SSR}{r}$	$\frac{\hat\beta_1^2S_{XX}}{\hat\sigma^2}=\frac{MSR}{MSE}$
Error	$SSE=\sum e^2=S_{YY}-SSR$	$n-1-r$	$\hat\sigma^2=\frac{SSE}{n-1-r}$	
Total	$SST=S_{YY}=\sum Y_i^2-n(\bar Y)^2$	$n-1$		

$$\hat\eta(x_0)=x_0^T\hat\beta\sim N(x_0^T\beta,\sigma^2x_0^T(X^TX)^{-1}x_0)\ \rightarrow\ \hat\eta(x_0)\pm\hat\sigma\sqrt{x_0^T(X^TX)^{-1}x_0}t_{(n-p-1),1-\frac{\alpha}{2}}$$

$$Y_0-x_0^T\hat\beta\sim N(0,\sigma^2[\mathbf{1}+x_0^T(X^TX)^{-1}x_0)\ \rightarrow\ x_0^T\beta\pm\hat\sigma\sqrt{\mathbf{1}+x_0^T(X^TX)^{-1}x_0}t_{(n-p-1),1-\frac{\alpha}{2}}$$

הראשון תוחלת והשני תחזית, כאשר $\hat\eta(x_0)=x_0^T\hat\beta$

רגרסיה לינארית פשוטה הנחות המודל: ε מתפלג נורמאלי עם σ^2.

$$Y=\beta_0+\beta_1 x+\varepsilon\ ,\qquad \varepsilon\sim N(0,\sigma^2)$$
$$E(Y|X=x)=\beta_0+\beta_1 x\ ,\qquad Var(Y|X=x)=\sigma^2$$

ה x-ים הם נתונים תמיד מבחינתי. צריך לאמוד את $\beta_0\beta_1\sigma^2$.

לכל x, ההתפלגות המותנה של Y בהינתן x היא נורמאלית.

השונות של Y בהינתן x קבועה ב x. **המשמעות** של σ^2 היא מידת הפיזור סביב הקו.

אומד ריבועים פחותים הוא הישר המביא למינימום את סכום ריבועי המרחקים האנכיים בין כל הנקודות לישר. (יודעים כאן את (X_i,Y_i) מהמדגם).

$$g(b_0,b_1)=\sum_{i=1}^{n}\left[\overbrace{Y_i-(b_0+b_1X_i)}^{\text{המרחק האנכי בין הנקודה ליישר}}\right]^2$$

האומדים ל β_0,β_1 הם אותם ערכי b_0,b_1 המביאים למינימום את g.

$$\frac{\partial}{\partial b_0}g(b_0,b_1)=-2\sum_{i=1}^{n}[Y_i-(b_0-b_1X_i)]\equiv 0$$
$$\frac{\partial}{\partial b_1}g(b_0,b_1)=-2\sum_{i=1}^{n}X_i[Y_i-(b_0-b_1X_i)]\equiv 0$$

ולכן נקבל את **המשוואות הנורמאליות**:

$$b_0\sum_{i=1}^{n}X_i+b_1\sum_{i=1}^{n}X_i^2=\sum_{i=1}^{n}X_iY_i \qquad nb_0+b_1\sum_{i=1}^{n}X_i=\sum_{i=1}^{n}Y_i$$

$$\hat\beta_0=b_0=\bar Y-\hat\beta_1\bar X$$
$$\hat\beta_1=b_1=\frac{\sum_{i=1}^{n}(X_i-\bar X)(Y_i-\bar Y)}{\sum_{i=1}^{n}(X_i-\bar X)^2}=\frac{S_{XY}}{S_{XX}}$$

$$S_{XY}=\sum(X_i-\bar X)(Y_i-\bar Y)=\sum(X_i-\bar X)Y_i=\sum X_i(Y_i-\bar Y)=\sum X_iY_i-n\bar X\bar Y$$
$$S_{XX}=\sum(X_i-\bar X)^2=\sum(X_i-\bar X)X_i=\sum X_i^2-n\bar X^2$$
$$S_{YY}=\sum(Y_i-\bar Y)^2=\sum(Y_i-\bar Y)Y_i=\sum Y_i^2-n\bar Y^2$$

- ניתן לאמוד את התוחלת המותנת של Y בהינתן x עבור x.
- ניתן לבנא עבור תצפית חדשה **בטווח**: $\hat Y=\hat\beta_0+\hat\beta_1 x$ (מחוץ: אקסטרפלציה).

נגדיר שגיאה $e_i=Y_i-\hat Y_i=Y_i-\hat\beta_0-\hat\beta_1X_i$ ונקבל:

$$\sum_{i=1}^{n}e_i=0\ ,\qquad \sum_{i=1}^{n}X_ie_i=0$$

$$E(\hat\beta_0)=\beta_0\ ,\qquad Var(\hat\beta_0)=\sigma^2\frac{\sum_{i=1}^{n}X_i^2}{nS_{XX}}$$

$$E(\hat\beta_1)=\beta_1\ ,\qquad Var(\hat\beta_1)=\sigma^2\frac{1}{S_{XX}}$$

$$\frac{(n-2)\hat\sigma^2}{\sigma^2}\sim\chi^2_{(n-2)}\ ,\qquad Cov(\hat\beta_0,\hat\beta_1)=-\sigma^2\frac{\bar X}{S_{XX}}$$

$$\hat\beta_0\sim N\left(\beta_0,\sigma^2\left[\frac{1}{n}+\frac{\bar X^2}{S_{XX}}\right]\right)\to\frac{\hat\beta_0-\beta_0}{\sqrt{\hat\sigma^2\left[\frac{1}{n}+\frac{\bar X^2}{S_{XX}}\right]}}\sim t_{(n-2)}$$

$$\hat\beta_1\sim N\left(\beta_1,\frac{\sigma^2}{S_{XX}}\right)\to\frac{\hat\beta_1-\beta_1}{\hat\sigma}\cdot\sqrt{S_{XX}}\sim t_{(n-2)}$$

אמידת השונות (SSE הן השוניות הלא מוסברות, המטרה היא להקטין אותו, SST קבוע).

$$SSE=\sum_{i=1}^{n}(e_i-\bar e)^2=\sum_{i=1}^{n}(Y_i-\hat Y_i)^2=\sum_{i=1}^{n}(Y_i-\hat\beta_0-\hat\beta_1x_i)^2=\sum_{i=1}^{n}e_i^2=(1-R^2)S_{YY}$$

$$SSR=\sum_{i=1}^{n}(\hat Y_i-\bar Y)^2=\sum_{i=1}^{n}\left(\hat\beta_0+\hat\beta_1x_i-(\hat\beta_0+\hat\beta_1\bar x)\right)^2=\hat\beta_1^2S_{XX}=\frac{S_{XY}^2}{S_{XX}}=R^2S_{YY}$$

$$SST=\sum_{i=1}^{n}(Y_i-\bar Y)^2=S_{YY}\ ,\qquad SST=SSR+SSE$$

$$\hat\sigma^2=\frac{SSE}{n-2}=\frac{\sum_{i=1}^{n}(Y_i-\hat Y_i)^2}{n-2}=\frac{1}{n-2}[S_{YY}-\hat\beta_1^2S_{XX}]\ ,\qquad E(\hat\sigma^2)=\sigma^2$$

$$\widehat{Var}(\hat\beta_0)=\hat\sigma^2\frac{\sum_{i=1}^{n}X_i^2}{nS_{XX}}\ ,\qquad \widehat{Var}(\hat\beta_1)=\hat\sigma^2\frac{1}{S_{XX}}$$

טענה: תחת הנחת הנורמאליות, $\hat\beta_1,\hat\beta_0$ ו $\hat\sigma^2$ הנם אומדי נראות מירבית. אנ"מ לשונות:

$$\check\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\hat\beta_0-\hat\beta_1X_i)^2=\frac{1}{n}\sum_{i=1}^{n}e_i^2$$

- $\hat\beta_1$ מתפלג נורמאלית (פונקציה ליניארית של מ"מ נורמאלית ב"ת).
- $\hat\beta_1,\hat\sigma^2$ ב"ת.

בדיקת השערות על **השיפוע**

$$H_0:\beta_1=\beta_1^0\ \text{vs}\ H_{1A}:\beta_1>\beta_1^0$$
$$H_0:\beta_1=\beta_1^0\ \text{vs}\ H_{1B}:\beta_1<\beta_1^0$$
$$H_0:\beta_1=\beta_1^0\ \text{vs}\ H_{1C}:\beta_1\neq\beta_1^0$$

סטטיסטי המבחן: $T=\dfrac{\hat\beta_1-\beta_1^0}{\hat\sigma/\sqrt{S_{XX}}}\underset{H_0}{\sim}t_{(n-2)}$

אם σ ידועה אז: $\dfrac{\hat\beta_1-\beta_1^0}{\sigma/\sqrt{S_{XX}}}\underset{H_0}{\sim}N(0,1)$

כלל הכרעה עבור מבחן ברמת מובהקות α:

A: $T>t_{(n-2),1-\alpha}$	A: $Pvalue=P(T\geq t)$
B: $T<-t_{(n-2),1-\alpha}$	B: $Pvalue=P(T\leq t)$
C: $T>t_{(n-2),1-\alpha/2}$ או $T<-t_{(n-2),1-\alpha/2}$	C: $Pvalue=P(\lvert T\rvert\geq\lvert t\rvert)$

ר"ס: $\hat\beta_1\pm\dfrac{\hat\sigma}{\sqrt{S_{XX}}}t_{(n-1),\left(1-\frac{\alpha}{2}\right)}$

בדיקת מובהקות הרגרסיה עבור $H_0:\beta_1=0$ (אחרת שונה מ0) היות ש $W\sim t_{(d)}\to W^2\sim F_{(1,d)}$ נקבל:

$$T^2=\left(\frac{\hat\beta_1}{\hat\sigma/\sqrt{S_{XX}}}\right)^2=\frac{\hat\beta_1S_{XX}}{\hat\sigma^2}=\frac{MSR}{MSE}\sim_{H_0}F_{(1,n-2)}$$

בדיקת השערות ור"ס **לחותך**: $T=\dfrac{\hat\beta_0-\beta_0^0}{\hat\sigma\sqrt{(nS_{XX})^{-1}\sum_{i=1}^{n}X_i^2}}\sim t_{(n-2)}$, $\hat\beta_0\pm\hat\sigma\sqrt{\dfrac{\sum_{i=1}^{n}X_i^2}{nS_{XX}}}t_{(n-1),\left(1-\frac{\alpha}{2}\right)}$

האומד לריבוע מקדם המתאם הלינארי מוגדר על ידי $R^2=\frac{SSR}{SST}=1-\frac{SSE}{SST}=\frac{\hat\beta_1^2S_{XX}}{S_{YY}}$

והוא אחוז השונות המוסברת של Y על ידי המודל. 0 מעיד על חוסר קשר לינארי 1 מעיד על קשר מלא (חיובי או שלילי).

אי-תלות המסקנה צריכה לגרום לשינוי הרגלים, אנחנו לעולם לא יודעים את **ההגורם**.

n תצפיות ב"ת.

כל תצפית מסווגת לפי **שני** משתנים:

הקטגוריות של משתנה אחד: $A_1,\dots,A_I$

הקטגוריות של משתנה שני: $B_1,\dots,B_J$

נסכם את הנתונים בטבלת שכיחויות דו-מימדית:

	B_1	B_2	$\cdots$	B_j	$\cdots$	B_J	סה"כ
A_1	X_{11}	X_{12}	$\cdots$	X_{1j}	$\cdots$	X_{1J}	$X_{1\cdot}$
A_2	X_{21}	X_{22}	$\cdots$	X_{2j}	$\cdots$	X_{2J}	$X_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	X_{i1}	X_{i2}	$\cdots$	X_{ij}	$\cdots$	X_{iJ}	$X_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_I	X_{I1}	X_{I2}	$\cdots$	X_{Ij}	$\cdots$	X_{IJ}	$X_{I\cdot}$
סה"כ	$X_{\cdot1}$	$X_{\cdot2}$	$\cdots$	$X_{\cdot j}$	$\cdots$	$X_{\cdot J}$	n

X_{ij} - השכיחות בתא ij

$X_{i\cdot}$ - השכיחות בשורה i

$X_{\cdot j}$ - השכיחות בעמודה j

המטרה: לבדוק האם יש קשר בין שני המשתנים.

ניסוח ההשערות:

$$H_0:P(A_i\cap B_j)=P(A_i)P(B_j)\qquad\forall ij$$
$$H_1:\ \text{otherwise}$$

לצורך בדיקת ההשערות לעיל עלינו למצוא את **השכיחות הצפויה** בכל אחת מהתאים שבטבלה **בהנחה שהשערת האפס נכונה**.

גודל המכונית

	גדולה	בינונית	קטנה
תאונה קטלנית	20	70	90
תאונה לא קטלנית	100	80	40

$$\hat P(A_1)=\frac{20+70+90}{400}=\frac{180}{400}=0.45\qquad \hat P(A_2)=\frac{100+80+40}{400}=\frac{220}{400}=0.55$$
$$\hat P(B_1)=\frac{130}{400}=0.325\qquad \hat P(B_2)=\frac{150}{400}=0.375\qquad \hat P(B_3)=\frac{120}{400}=0.3$$

לכן, אם המשתנים ב"ת נציע

$$\hat P(A_1\cap B_1)=0.45\cdot0.325=0.14625\qquad \hat P(A_1\cap B_2)=.45\cdot0.375=0.16875\qquad \hat P(A_1\cap B_3)=0.45\cdot0.3=0.135$$
$$\hat P(A_2\cap B_1)=0.55\cdot0.325=0.17875\qquad \hat P(A_2\cap B_2)=.55\cdot0.375=0.20625\qquad \hat P(A_2\cap B_3)=0.55\cdot0.3=0.165$$
$$\sum_{i=1}^{I}\sum_{j=1}^{J}\hat P(A_i\cap B_j)=1$$

$$\hat P(A_1\cap B_1)=0.14625\qquad \hat P(A_1\cap B_2)=0.16875\qquad \hat P(A_1\cap B_3)=0.135$$
$$\hat P(A_2\cap B_1)=0.17875\qquad \hat P(A_2\cap B_2)=0.20625\qquad \hat P(A_2\cap B_3)=0.165$$

ולכן טבלת **השכיחויות הצפויות** הנה

גודל המכונית

	גדולה	בינונית	קטנה
תאונה קטלנית	400·0.135	400·0.16875	400·0.14625
תאונה לא קטלנית	400·0.165	400·0.20625	400·0.17875

סכום התאים בטבלה: 400

באופן כללי:

$$\hat P(A_i)=\frac{X_i}{n}\ ,i=1,\dots,I$$
$$\hat P(B_j)=\frac{X_j}{n}\ ,j=1,..,J$$

לכן אם המשתנים ב"ת

$$\hat P_{H_0}(A_i\cap B_j)=\frac{X_i}{n}\cdot\frac{X_j}{n}\ \forall ij$$

והשכיחות הצפויה בתא ה ij היא:

$$E_{H_0}(X_{ij})=\frac{X_iX_j}{n}$$

סטטיסטי המבחן לבדיקת אי-תלות מוגדר להיות:

$$X^2=\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(X_{ij}-\frac{X_iX_j}{n}\right)^2}{\frac{X_iX_j}{n}}$$

ועבור מדגם מספיק גדול, אם השערת האפס נכונה אז:

$$X^2\sim\chi^2_{(I-1)(J-1)}$$

מקובל להניח שהקירוב מספיק טוב אם:

$$\frac{X_iX_j}{n}\geq 5,\forall ij$$

הסבר לדרגות החופש:

$$\underbrace{IJ-1}_{\text{סכום תאי הטבלה שווה לגודל המדגם}}-\underbrace{(I-1)}_{\text{אמידת }\hat P(A_i)}-\underbrace{(J-1)}_{\text{אמידת }\hat P(B_j)}=$$
$$=IJ-I-J+1=$$
$$=(I-1)(J-1)$$

$$\begin{pmatrix}a&b\\c&d\end{pmatrix}^{-1}=\frac{1}{ad-bc}\begin{pmatrix}d&-c\\-b&a\end{pmatrix}$$

$$\begin{pmatrix}a&b&c\\b&d&e\\c&e&f\end{pmatrix}^{-1}=$$
$$=\frac{1}{\Delta}\begin{pmatrix}df-e^2&ce-bf&be-cd\\ce-bf&af-c^2&bc-ae\\be-cd&bc-ae&ad-b^2\end{pmatrix}$$
$$\Delta=adf+2bce-ae^2-b^2f-c^2d$$

רגרסיה לינארית מרובה עבור p משתנים מסבירים:

$$Y_i=\beta_0+\beta_1X_{1i}+\cdots\beta_pX_{pi}+\varepsilon_i\ ,\qquad \varepsilon_i\sim N(0,\sigma^2)\ iid, i=1,\dots,n$$
$$\vec Y=X_{n\times p}\vec\beta+\vec\varepsilon\ ,\qquad \vec\varepsilon\sim MVN(0,\sigma^2I)$$

התפלגות רב-נורמאלית כאשר $\vec\mu\in R^m$ ו $V_{m\times m}$ חיובית לחלוטין $W\sim MVN(\vec\mu,V)$.

$$f_W(w)=(2\pi)^{-\frac{m}{2}}|V|^{-\frac{1}{2}}\cdot e^{-\frac{1}{2}(\vec w-\vec\mu)^TV^{-1}(\vec w-\vec\mu)}$$

אר"פ יוגדרו להיות ערכי המקדמים המביאים למינימום את:

$$g(b_0,b_1,\dots,b_p)=\sum_{i=1}^{n}[Y_i-(b_0+b_1X_{1i}+\cdots+b_pX_{pi}]^2=(Y-Xb)^T(Y-Xb)=Y^TY-2b^TX^TY+b^TX^TXb$$

אזי נגזור ונקבל $p+1$ משוואות נורמאליות $X^TXb=X^TY$.

אם חלק מהמשוואות תלויות ליניארית (יש תלות בין ה X-ים) אז X^TX איננה הפיכה. אחרת:

$$\hat\beta=(X^TX)^{-1}X^TY\ ,\ \hat Y=X\hat\beta=X(X^TX)^{-1}X^TY\ ,\ e=Y-\hat Y$$

$$X^Te=0\ ,\qquad \sum_{i=1}^{n}e_i=0\ ,\qquad \sum_{i=1}^{n}e_i\hat Y_i=e^T\hat Y=0$$

$$\overbrace{\sum_{i=1}^{n}(Y_i-\bar Y)^2}^{SST=S_{YY}}=\overbrace{\sum_{i=1}^{n}(\hat Y_i-\bar Y)^2}^{SSR}+\overbrace{\sum_{i=1}^{n}(Y_i-\hat Y_i)^2}^{SSE}\ ,\qquad R^2=\frac{SSR}{SST}$$

$$SSE=||e||^2\ ,\qquad SSR=Y^TX\hat\beta-n(\bar Y)^2=\hat\beta^TX^TY-n(\bar Y)^2$$
$$E(AZ)=AE(Z)\ ,\qquad Var(AZ)=ACov(Z)A^T$$
$$E(\hat\beta)=\beta\ ,\qquad Var(\hat\beta)=\sigma^2(X^TX)^{-1}$$

חסר הטיה

$$\hat{\sigma^2}=\frac{1}{n-p-1}\sum_{i=1}^{n}e_i^2=\frac{e^Te}{n-p-1}=\frac{S_{YY}-SSR}{n-p-1}\to E(\hat\sigma^2)=\sigma^2$$

$$\hat\beta\sim MVN(\beta,\sigma^2(X^TX)^{-1})\ ,\qquad \hat Y\sim MVN(X\beta,\sigma^2X(X^TX)^{-1}X^T)$$
$$e\sim MVN(0,\sigma^2(I-X(X^TX)^{-1}X^T))\ ,\qquad \frac{n-p-1}{\sigma^2}\hat\sigma^2=\frac{e^Te}{\sigma^2}\sim\chi^2_{(n-p-1)}$$

מטריצת השונות המשותפת של $\hat\beta$ היא $\hat\sigma^2(X^TX)^{-1}$ ולכן $Var(\hat\beta_i)=\hat\sigma^2a_{ii},Cov(\hat\beta_i,\hat\beta_j)=\hat\sigma^2a_{ij}$

$$(X^TX)=\begin{pmatrix}n&\sum X_i\\\sum X_i&\sum X_i^2\end{pmatrix}=\begin{pmatrix}n&\sum X_{1i}&\sum X_{2i}\\\sum X_{1i}&\sum X_{1i}^2&\sum X_{1i}X_{2i}\\\sum X_{2i}&\sum X_{1i}X_{2i}&\sum X_{2i}^2\end{pmatrix}$$
$$(X^TY)=\begin{pmatrix}\sum Y_i\\\sum X_{1i}Y_i\\\sum X_{2i}Y_i\end{pmatrix}$$

רווח סמך לכל מקדם בנפרד. סטטיסטי המבחן (עבור 0 המשמעות היא תרומת המקדם למודל).

$$\frac{\hat\beta_j^0-\beta_j^0}{\hat\sigma\sqrt{(X^TX)^{-1}_{(j+1),(j+1)}}}\sim_{H_0}t_{(n-p-1)}\to\ \hat\beta_j\pm\hat\sigma\sqrt{(X^TX)^{-1}_{(j+1),(j+1)}}\cdot t_{(n-p-1),1-\frac{\alpha}{2}}$$

בדיקת מובהקות הרגרסיה. $\vec\beta=\vec0$ דוחים אם:

$$\frac{\frac{SSR}{p}}{\hat\sigma^2}=\frac{\frac{SSR}{p}}{\frac{e^Te}{n-p-1}}>F^{1-\alpha}_{(p,n-p-1)}$$

בדיקת השערה של מובהקות חלק ממקדמי הרגרסיה. $\beta_{m+1}=\beta_{m+2}=\cdots=\beta_p=0$ דוחים אם:

$$\frac{\frac{SSR(1,\dots,p)-SSR(1,\dots,m)}{p-m}}{\hat\sigma^2(1,\dots p)}F^{1-\alpha}_{(p-m,n-p-1)}$$

Stepwise Regression Methods לא בהכרח מובילות למודל הטוב ביותר.

Forward – הוספת משתנה אחת בכל צעד. תחיל בוחרים את הממקסם של SSR ובכל צעד את המינימום של SSE.

עד שהמשתנה האחרון אינו מובהק ואותו לא מכניסים.

$$PartialSS=SSR(\beta_{new}|\beta_{old})=SSR(\beta_{new},\beta_{old})-SSR(\beta_{old})$$

Backward – מתחילים מכל המודלים ומוצאים את מי שמעלה את ה SSE הכי מעט. עוצרים אם כל המשתנים **שנותרנו** במודל מובהקים ברמת מובהקות שנקבעה מראש.

(האלגוריתמים שונים ולכן נקבל תוצאות שונות גם עבור רמות מובהקות שוות).

Stepwise – מספר חישובים רב יותר, שתי רמות מובהקות: להיכנס ולצאת. עוצרים אם מבין כל המשתנים **שאינם** במודל, אין משתנה העובר את קריטריון הכניסה. **המודל הסופי הוא המודל שהתקבל בשלב האחרון.**