Hinweise zur 13.Übung: α- und β-Fehler, KS-Test und MW-Vergleiche

1.] Wir stellen uns folgende Situation vor: Eine Münze sei entweder "ehrlich", d.h. $P(K) = P(W) = \frac{1}{2}$, oder gewichtsmäßig unsymmetrisch mit $P(K) = \frac{3}{10}$, und $P(W) = \frac{7}{10}$. Durch n Stichproben für $P(K)$ soll dies geklärt werden. Dabei sei als kritischer Wert δ_k=0.4 gewählt, der Mittelwert von $\frac{3}{10}$ und $\frac{1}{2}$. Wie groß muß der Stichprobenumfang n sein, um bei $\delta_k \cdot n$ als kritische Grenze der Stichprobe eine 95%-ige Sicherheit gegen den α-Fehler zu haben? Ist dies wegen der Symmetrie von δ_k dann auch die analoge Grenze des β-Fehlers nach unten? Begründen Sie die Aussage und versuchen Sie eine Lösung mit SPSS!

Wenn der Stichprobenumfang n größer wird, wird die Varianz enger, und die 5%-Grenze der Nullhypothese rückt näher an 0.5 heran. Folgende Befehlsdatei ist ein erster Test für $n = 101$.

```
/* Aktiviere 101 Zeilen */
Input Program .
LOOP #I=1 to 101 .
Compute anzahl=#I .
FORMATS anzahl (F8).
END CASE .
END LOOP .
END FILE .
END INPUT PROGRAM .
EXECUTE .
/* Verteilungen und Dichten der beiden Faelle */
COMPUTE bi03 = CDF.BINOM($casenum,101,0.3) .
EXECUTE .
COMPUTE bi05 = CDF.BINOM($casenum,101,0.5) .
EXECUTE .
CREATE
 /bi03D=DIFF(bi03 1) /bi05D=DIFF(bi05 1).
GRAPH
 /LINE(MULTIPLE)= VALUE( bi03D bi05D ) .
```

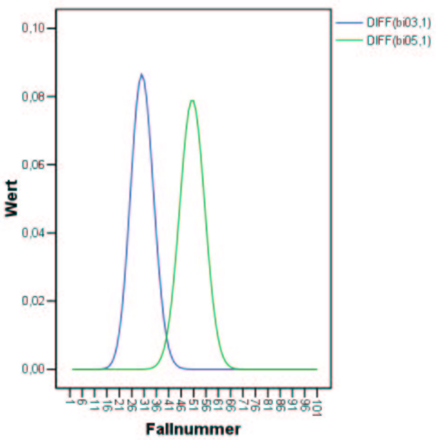


Figure 1: Dichten bi03D und bi05D fuer n=101

Im DatenFenster kann man in der Spalte der Verteilungsfunktion bi05 der "ehrlichen" Münze ablesen, ob für $\delta_k \cdot n = [0{,}4 \cdot 101] + 1 = 41$ Versuche der Wert 5% überschritten ist: Es ist noch weniger als 2%, also war n zu groß angesetzt. (Wenn man es sich leisten kann: o.k., aber wenn die Versuche teuer sind, sollte man das kleinst-mögliche n suchen. Auch bei der Ablehnungsquote vergibt man sich etwas.) Die Linienplots geben zusätzlich einen Eindruck der beteiligten Dichten.
Iterativ kann obige Befehlsfolge nun so oft für verschiedene n probiert werden, bis wir exakt die gesuchte Grenze erreichen. Dabei tritt noch eine Komplikation auf: Die Anzahl der Versuche ist ganzzahlig, und wir sollten eine ganzzahlige Grenze für den kritischen Wert angeben. Der ganze Anteil $[\delta_k \cdot n]$ aber springt.
Beispiel:

```
/* De-Aktiviere  Zeilen 71-101, d.h. einfach loeschen  */
COMPUTE bi03_70 = CDF.BINOM($casenum,70,0.3) .
EXECUTE .
COMPUTE bi05_70 = CDF.BINOM($casenum,70,0.5) .
EXECUTE .
CREATE
 /bi05_70D=DIFF(bi05_70 1) /bi03_70D=DIFF(bi03_70 1).
GRAPH
 /LINE(MULTIPLE)= VALUE( bi05_70D bi03_70D ) .
```

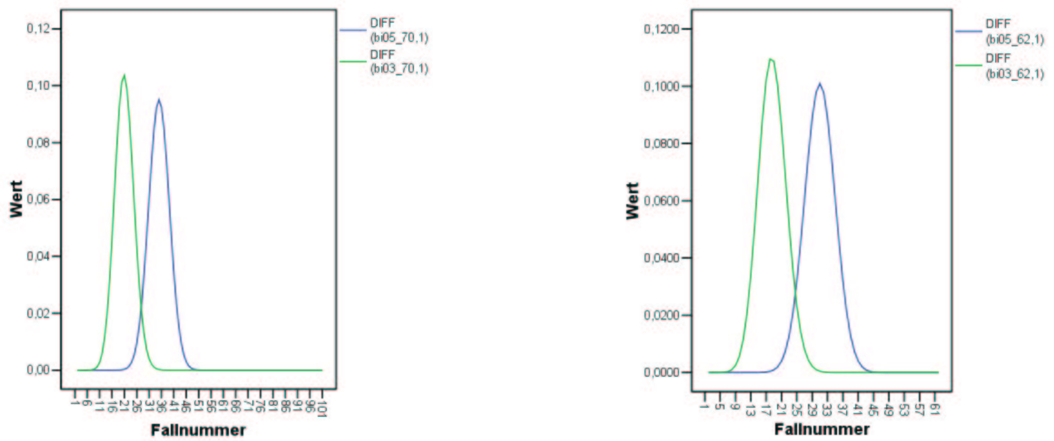


Figure 2: Dichten fuer n=70, und n=62

Es zeigt sich, dass n noch zu groß ist: $k_c = 28$ ist mit 6% genau 1 zu groß, bei 27 haben wir 3,6%. Wegen der Ganzzahligkeit erhält man die beste Lösung für n=62.

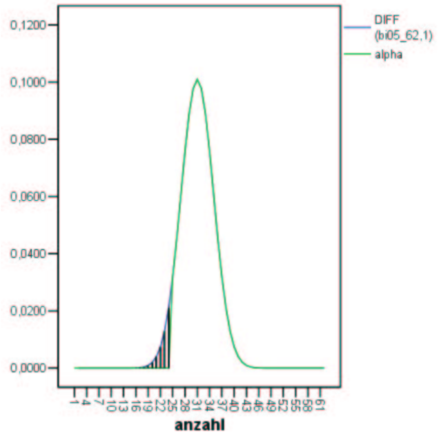
```
/* De-Aktiviere  Zeilen 63-71, d.h. einfach loeschen  */
COMPUTE bi05_62 = CDF.BINOM($casenum,62,0.5) .
EXECUTE .
COMPUTE bi03_62 = CDF.BINOM($casenum,62,0.3) .
EXECUTE .
CREATE
 /bi05_62D=DIFF(bi05_62 1) /bi03_62D=DIFF(bi03_62 1).
GRAPH
 /LINE(MULTIPLE)= VALUE( bi05_62D bi03_62D ) .
```

Beim Wert k_c=[0,4*n]+1=25 ist in der Datentabelle von bi05_62 nachzusehen, ob die Verteilungsfunktion 0,05 ueberschritten hat; bei 24 finden wir 4,9%, bei 25 dann

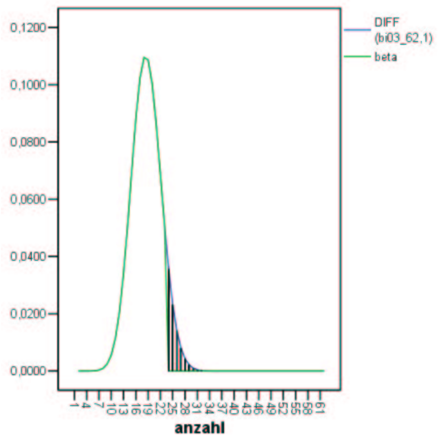
8,1%. Also sollte der Stichprobenumfang 62 sein, und der kritische Wert einer Strichprobe zur Ablehnung der Nullhypothese ist dann 24.
Die Streuungen der beiden Hypothesen sind nicht gleich:

$$0.25/n = p_0(1-p_0)/n \neq p_A(1-p_A)/n = 0.21/n\ ,$$

also liegt $k_c * n$ nicht in der "Mitte" bezüglich α- und β-Fehler. Bei $n = 62$ und k_c=24 ist α=0.05 erreicht, aber der β-Fehler ist etwa 5,4%. Dies ist trotzdem noch recht ausgewogen.
α-Fehler:



β-Fehler:



2.] a) Erzeugen Sie mit Hilfe der Funktion RV.NORMAL 222 nach N(0, 1000) normalverteilte Zufallszahlen als Variable NN.
b) Zeichnen Sie mit Hilfe von STREUDIAGRAMM die empirische Verteilungsfunktion von NN, sowie die theoretische Verteilungsfunktion von N(0, 1000), und bilden Sie die Differenz beider Funktionen.
c) Vergleichen Sie das Resultat mit dem Kolmogorov-Smirnov Test für NN.

Man aktiviere 222 Zeilen und erzeuge NN=RV.NORMAL(NN,0,1000) . Durch Sortieren von NN kann man mit EmpVert = $casenum /222 die empirische Verteilungsfunktion erzeugen. Der zu jedem NN gehörige Wert der "wahren" Verteilungsfunktion ergibt sich durch direktes → Berechnen als CDF-Funktion Vert = CDF.NORMAL(NN,0,1000). (Man kann auch genauer die in SPSS berechenbaren realen Werte für Mittelwert und Streuung von NN in die CDF einssetzen.) Die Test-Statistik des KS-Tests ist dif=$\sqrt{222}\, max|Vert - EmpVert|$ =0,72566 . In → Nichtlineare Tests, KS-Test, ergibt

sich hier ein Signifikanzniveau von über 96%, also wird die Zufallsverteilung von SPSS auch als Normalverteilung ausgewiesen.
Der Test auf Normalverteilung soll ausschließen, daß wir nichtnormal verteilte Daten fälschlich als solche ansehen: Wir brauchen also ein kleines "Konsumentenrisiko", also einen kleinen β-Fehler. Man geht mit einer $\boxed{\text{groben Faustregel}}$ davon aus, daß für ein Signifikanzniveau von $\boxed{\geq 40\%}$ der β-Fehler klein genug ist. Dies ist hier auf jeden Fall mit 96% ausgewiesen.

Die nächsten Aufgaben betreffen verschiedene Mittelwertvergleiche.
Bei abhängigen (gepaarten) Stichproben und

bei normalverteilten Merkmalen verwende man

$\rightarrow$ Mittelwertvergleiche $\rightarrow$ gepaarter T-Test

bei nicht normalverteilten Merkmalen verwende man

$\rightarrow$ Nichtparametrische Tests $\rightarrow$ Wilcoxon Test

Bei unabhängigen Stichproben und

bei normalverteilten Merkmalen verwende man

$\rightarrow$ Mittelwertvergleiche $\rightarrow$ T-Test für unabhängige Stichproben

bei nicht normalverteilten Merkmalen verwende man

$\rightarrow$ Nichtparametrische Tests $\rightarrow$ Mann-Whitney Test

3.] Laden Sie die Daten von Z:rheuma.sav . Es soll geprüft werden, ob die Merkmale für α2-Globulin Lc10$_1$ und Lc10$_2$ Mittelwertunterschiede aufweisen. Anleitung: Zuerst ist zu prüfen, ob beide Merkmale normalverteilt sind (KS-Test). Nutzen Sie dabei die Option $\rightarrow$ Fallweiser Ausschluß. Für den Test selbst verwende man den T-Test bei gepaarten Stichproben. Die Antwort soll zu einem Signifikanzniveau α=0.05 erfolgen.

Beim KS-Test ist bei $\rightarrow$ Optionen $\rightarrow$ Fallweiser Ausschluß anzuklicken. Damit soll gesichert werden, daß nicht Datenzeilen mit fehlenden einzelnen Werten den Test verfälschen. Der KS-Test ergibt die beiden Werte (für 2-Tailed P) von 0,334 und 0,605. Der erstere ist nicht ganz 0,4, aber man gehe davon aus, daß trotzdem noch näherungsweise eine Normalverteilung vorliegt.
Der T-Test für den Mittelwertvergleich zweier gepaarter Stichproben verwendet die direkten Differenzen d_i aller einzelnem Ausprägungen für alle n besetzten Zeilen. Die Test-Statistik ist dann

$$t = \sqrt{n}\frac{\bar{d}}{s_d}$$

einer t-Verteilung mit $n-1$ Freiheitsgraden. Wenn $|t| > t_{n-1;1-\alpha/2}$ ist, ist die Nullhypothese abzulehnen.
Im T-Test für den Mittelwertvergleich von Lc10$_1$ und Lc10$_2$ erhalten wir nur die 2-tail Sig(nifikanz) 0,009, d.h. weniger als 1%, was für eine Bestätigung der Nullhypothese (gleiche Mitterwerte beider Stichproben) zuwenig ist. Dies kann man auch an dem mit ausgegebenen CI-Intervall für die Differenz beider Mittelwerte ablesen: Der Nullpunkt ist nicht mit eingeschlossen.